

# Interactive Schemes for the AWGN Channel with Noisy Feedback

Assaf Ben-Yishai and Ofer Shayevitz

**Abstract**—We study the problem of communication over an additive white Gaussian noise (AWGN) channel with an AWGN feedback channel. When the feedback channel is noiseless, the classic Schalkwijk-Kailath (S-K) scheme is known to achieve capacity in a simple sequential fashion, while attaining reliability superior to non-feedback schemes. In this work, we show how simplicity and reliability can be attained even when the feedback is noisy, provided that the feedback channel is sufficiently better than the feedforward channel. Specifically, we introduce a low-complexity low-delay interactive scheme that operates close to capacity for a fixed bit error probability (e.g.  $10^{-6}$ ). We then build on this scheme to provide two asymptotic constructions, one based on high dimensional lattices, and the other based on concatenated coding, that admit an error exponent significantly exceeding the best possible non-feedback exponent. Our approach is based on the interpretation of feedback transmission as a side-information problem, and employs an interactive modulo-lattice solution.

## I. INTRODUCTION

While feedback cannot increase the capacity of point-to-point memoryless channels [1], there exist noiseless feedback communication schemes that can provide a significant improvement in terms of simplicity and reliability, see e.g. [2]–[5]. However, these elegant feedback schemes completely fail in the presence of arbitrarily low feedback noise, rendering them grossly impractical. This naturally raises the question of whether simplicity and reliability can still be achieved to some degree in a practical setup of noisy feedback. In this paper, we address this question in a Gaussian setting and answer it in the affirmative.

The setup we consider is the following. Two Terminals A and B are connected by a pair of independent AWGN channels, and are limited by individual power constraints. The channel from Terminal A (resp. B) to Terminal B (resp. A) is referred to as the feedforward (resp. feedback) channel. Terminal A is in possession of a message to be reliably transmitted to Terminal B. To that end, an interactive communication model

is adopted where both terminals are allowed to employ coding and exchange signals on the fly. This model is sometimes referred to as *active feedback*, and should be distinguished from the *passive feedback* setting where no coding is allowed over the feedback channel. The AWGN channel with noiseless feedback was studied in the classical works of Schalkwijk and Kailath [2], [3], who introduced a capacity-achieving communication scheme referred to herein as the *S-K Scheme*. This linear-feedback coding scheme employs a first-order recursion at both terminals, and is markedly simpler than its non-feedback counterparts that typically employ long block codes and complex encoding/decoding techniques. In terms of reliability, the error probability attained by the S-K scheme decays super-exponentially with the delay, in contrast to the weaker exponential decay achieved by non-feedback codes [6]. However, this scheme and its linear feedback generalizations are not robust to any amount of feedback noise, as was initially observed in [3] and further strengthened in [7].

The main contribution of this work is in showing that to some extent, the merits of noiseless feedback can be carried over to the practical regime of noisy feedback. Contrary to the noiseless feedback case, these improvements in simplicity and reliability are not simultaneously achieved. In terms of reliability, we construct two interactive protocols that are of comparable complexity to non-feedback schemes, but are superior in the asymptotic error exponent sense. In terms of simplicity, we depart from the standard asymptotic regime and show how a fixed (but low) error probability can be attained at a small capacity gap, where the latter term refers to the amount of excess SNR required by the scheme above the minimum predicted by the Shannon limit. Both these constructions are useful when the signal-to-noise ratio of the feedback channel sufficiently exceeds that of the feedforward channel.

As a case in point, consider the high-SNR regime and assume that the SNR of the feedback channel exceeds the SNR of the feedforward channel by 20dB. Then our simplicity-oriented scheme operates at a capacity gap of merely 0.8dB with only 19 rounds of interaction, and attains a bit error rate of  $10^{-6}$ . This should be juxtaposed against two reference systems, operating at the same bit error rate: On the one hand, state-of-the-art non-feedback codes that attain the same

The authors are with the Department of EE-Systems, Tel Aviv University, Tel Aviv, Israel {assafbster@gmail.com, ofersha@eng.tau.ac.il}. The work of A. Ben-Yishai was partially supported by an ISF grant no. 1367/14. The work of O. Shayevitz was supported by an ERC grant no. 639573, a CIG grant no. 631983, and an ISF grant no. 1367/14. This paper was presented in part at Allerton 2014 and ISIT 2015.

capacity gap require roughly two orders of magnitude increase in delay and complexity. On the other hand, the capacity gap attained by a minimal delay uncoded system is at least 9dB. Finally, under the same setup, our reliability-oriented schemes attain an error exponent exceeding the sphere-packing bound of the feedforward channel for a wide range of rates below capacity.

The construction we introduce is based on endowing the S-K scheme with modulo-lattice operations. As observed in [8], the feedforward S-K scheme can be interpreted as a solution to a *Joint Source-Channel Coding* (JSCC) problem via analog transmission. Here, we further observe that the feedback transmission problem can be cast as a similar problem but with side information (i.e. the message) at the receiver (i.e., Terminal A). This observation is crucial for our construction, and is leveraged by means of modulo-lattice analog transmission in the spirit of Kochman and Zamir [9].

Let us briefly describe the simplicity-oriented version of our scheme. Terminal A encodes its message into a scalar  $\Theta$  using pulse amplitude modulation (PAM). In subsequent rounds, Terminal B computes a linear estimate of  $\Theta$ , and feeds back an exponentially amplified version of this estimate, modulo a fixed interval. The modulo operation facilitates the essential “zoom-in” amplification without exceeding the power limit, at the cost of a possible modulo-aliasing error. In turn, Terminal A employs a suitable modulo computation and obtains (if no modulo-aliasing occurs) the estimation error, corrupted by excess additive noise. This quantity is then properly scaled and sent over the feedforward channel to Terminal B. After a fixed number of rounds, Terminal B decodes the message using a minimum distance rule. Loosely speaking, the scheme’s error probability is dictated by the events of a modulo-aliasing in any of the rounds, as well as the event where the remaining estimation noise is larger than the minimum distance of the PAM.

We also introduce two asymptotic reliability-oriented versions of our scheme. The first is based on an asymptotic generalization of the simple interaction idea above, where a block code replaces the PAM modulation, and a block S-K scheme is used in conjunction with a multi-dimensional modulo-lattice operation, replacing the scalar modulo. We provide an asymptotic error exponent analysis using the Poltyrev exponent to account for modulo-aliasing errors, and channel coding exponents to account for the error of the block code. The second scheme we present is based on concatenated coding, with the scalar simplicity-oriented scheme as an inner code and a block outer code. Since the discrete memoryless channel (DMC) induced by the inner code and viewed by the outer code is non-Gaussian, we give a lower bound for the error exponent based on the performance of a “worst case”

symmetric DMC.

*Related work.* In [7], [10], the authors analyzed the reliability function of the AWGN channel at zero rate for noisy *passive* feedback, i.e. where the channel outputs are fed back without any processing. In [11], the authors gave an interesting analysis of the reliability of transmission of an  $M$ -ary message ( $M \geq 3$ ) over AWGN with passive noisy feedback, but in a slightly different setting where a peak energy constraint is imposed. In [12], the authors considered a concatenated coding scheme with a passive linear-feedback inner code and a block outer code, and provided some error exponent results. In Section VI-G, we compare our reliability-oriented scheme to [12], and show that the exponent obtained in [12] is better for low rates whereas our exponent is better for high rates. In [13], which is closer to our interactive setting, the reliability function associated with the transmission of a single bit over an AWGN channel with noisy *active* feedback has been considered. Specifically, it was shown that active feedback roughly quadruples the error exponent relative to passive feedback. The achievability result of [13] is better than ours at zero rate but does not extend to positive rates.

*Organization.* Notation and definitions are given in Section II. The problem setup is introduced in Section III. Necessary background is given in Section IV. Simple interaction is addressed in Section V, and improving reliability is addressed in Section VI.

## II. NOTATION AND DEFINITIONS

In the sequel, we use the following notation. For any number  $x > 0$ , we write  $x_{\text{dB}} \stackrel{\text{def}}{=} 10 \log_{10}(x)$  to denote the value of  $x$  in decibels. The Gaussian Q-function is

$$Q(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-u^2/2) du$$

and  $Q^{-1}(\cdot)$  is its functional inverse. We write  $f(x) = O(g(x))$  for  $\limsup_{x \rightarrow \infty} |f(x)/g(x)| < \infty$ . We write  $\log$  for base 2 logarithm, and  $\ln$  for the natural logarithm. We use the vector notation  $\mathbf{x}^n \stackrel{\text{def}}{=} (x_1, \dots, x_n)$  and boldface letters such as  $\mathbf{x}$  to indicate vectors of size  $N_\Lambda$ . We write  $a_n \geq b_n$  to mean  $\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \left( \frac{a_n}{b_n} \right) \geq 0$ , and similarly define  $\leq$  and  $\doteq$ . We use  $\bar{A}$  to denote the complementary of an event  $A$ .

*The Capacity Gap.* Recall that the Shannon capacity of the AWGN channel with signal-to-noise ratio SNR is given by

$$C = \frac{1}{2} \log(1 + \text{SNR}). \quad (1)$$

This is the maximal rate achievable by any scheme (of unbounded complexity/delay, with or without feedback) under vanishing error probability. Conversely, the minimal SNR required to reliably attain a rate  $R$  is  $2^{2R} - 1$ . The *capacity gap*  $\Gamma$  attained by a coding scheme that operates at rate  $R$  over

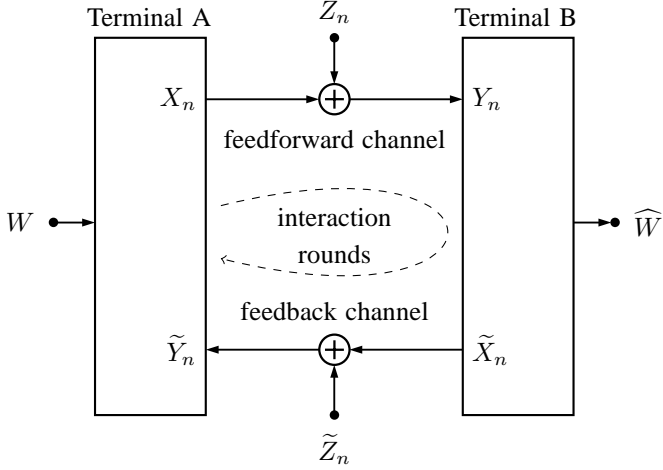


Fig. 1. Block diagram of interactive coding over an AWGN channel with noisy feedback

an AWGN channel, is the excess SNR required by the scheme over the minimum predicted by the Shannon limit, i.e.,

$$\Gamma \stackrel{\text{def}}{=} \frac{\text{SNR}}{2^{2R} - 1}. \quad (2)$$

Note that if a nonzero bit/symbol error probability is allowed, then one can achieve rates exceeding the Shannon capacity (1), and this effect should in principle be accounted for, to make the definition of the capacity gap fair. However, for small error probabilities the associated correction factor (related to the corresponding rate-distortion function) becomes negligible, and we therefore ignore it in the sequel.

### III. SETUP

Our problem setup is depicted in Fig. 1. The feedforward and feedback channels connecting Terminal A to Terminal B and vice versa, are AWGN channels given by

$$\begin{aligned} Y_n &= X_n + Z_n, \\ \tilde{Y}_n &= \tilde{X}_n + \tilde{Z}_n, \end{aligned}$$

where  $X_n, Y_n$  (resp.  $\tilde{X}_n, \tilde{Y}_n$ ) are the input and output of the feedforward (resp. feedback) channel at time  $n$  respectively. The feedforward (resp. feedback) channel noise  $Z_n \sim \mathcal{N}(0, \sigma^2)$  (resp.  $\tilde{Z}_n \sim \mathcal{N}(0, \tilde{\sigma}^2)$ ) is independent of the input  $X_n$  (resp.  $\tilde{X}_n$ ), and constitutes an i.i.d. sequence. The feedforward and feedback noise processes are mutually independent.

Terminal A is in possession of a message  $W$ , uniformly distributed over the set  $\{0, \dots, M-1\}$ , to be described to Terminal B over  $N$  rounds of communication. To that end, the terminals can employ an interactive scheme defined by the sequences of functions  $\varphi_n$  and  $\tilde{\varphi}_n$  as follows: At time  $n$ , Terminal A sends a function of its message  $W$  and possibly of

past feedback channel outputs over the feedforward channel, i.e.,

$$X_n = \varphi_n(W, \tilde{Y}^{n-1}).$$

Similarly, Terminal B sends function of its past observations to Terminal A over the feedback channel, i.e.,

$$\tilde{X}_n = \tilde{\varphi}_n(Y^n).$$

*Remark 1:* In general, we allow these functions to further depend on common randomness shared by the terminals. We note in passing that our definition of the feedback transmission scheme is sometimes referred to as *active feedback*; the term *passive feedback* is usually reserved to the special case where  $\tilde{\varphi}(Y^n) = Y^n$ .

The number of rounds  $N$  is fixed. While feedback protocols with variable transmission length exist and can improve reliability relative to non-feedback transmission [16], [17], they are beyond the scope of this work. We assume that Terminal A (resp. Terminal B) is subject to an average power constraint  $P$  (resp.  $\tilde{P}$ ), namely

$$\sum_{n=1}^N \mathbb{E}(X_n^2) \leq N \cdot P, \quad \sum_{n=1}^N \mathbb{E}(\tilde{X}_n^2) \leq N \cdot \tilde{P}.$$

We denote the feedforward (resp. feedback) signal-to-noise ratio by  $\text{SNR} \stackrel{\text{def}}{=} \frac{P}{\sigma^2}$  (resp.  $\tilde{\text{SNR}} \stackrel{\text{def}}{=} \frac{\tilde{P}}{\tilde{\sigma}^2}$ ). The excess signal-to-noise ratio of the feedback over the feedforward is denoted by  $\Delta\text{SNR} \stackrel{\text{def}}{=} \frac{\tilde{\text{SNR}}}{\text{SNR}}$ . Throughout this work, we assume that  $\Delta\text{SNR} > 1$ .

An interactive scheme  $(\varphi, \tilde{\varphi})$  is associated with a rate  $R \stackrel{\text{def}}{=} \frac{\log M}{N}$  and an error probability  $p_e$ , which is the probability that Terminal B errs in decoding the message  $W$  at time  $N$ , under the optimal decision rule.

### IV. PRELIMINARIES

In this section, we describe the building blocks underlying our interactive scheme. First, we review the use of uncoded PAM signaling, and discuss its associated capacity gap. Then, we describe the basic problem of joint source-channel coding (JSCC) via analog transmission, and show how to build the S-K scheme from uncoded PAM and iterative JSCC. Lastly, we discuss the problem of JSCC with side information using modulo arithmetic, and present a simple scalar solution that is later implemented as part of our simplicity-oriented scheme.

#### A. Uncoded PAM

PAM is a simple and commonly used modulation scheme, where  $2^R$  symbols are mapped (one-to-one) to the set

$$\{\pm 1\eta, \pm 3\eta, \dots, \pm(2^R - 1)\eta\}.$$

Canonically, the normalization factor  $\eta$  is set so that the overall mean square of the constellation (assuming equiprobable symbols) is unity. A straightforward calculation yields  $\eta = \sqrt{3/(2^{2R}-1)}$ . In the general case where the mean square of the constellation is constrained to be  $P$ ,  $\eta$  is replaced with  $\eta\sqrt{P}$ .

It is easy to show that for an AWGN channel with zero mean noise of variance  $\sigma^2$  and average input power constraint  $P$ , the probability of error incurred by the optimal detector is given by the following formula

$$\begin{aligned} p_e &= 2(1 - 2^{-R}) Q\left(\frac{\sqrt{P}\eta}{\sigma}\right) \\ &= 2(1 - 2^{-R}) Q\left(\sqrt{\frac{3\text{SNR}}{2^{2R}-1}}\right). \end{aligned} \quad (3)$$

Manipulating (3) yields:

$$R = \frac{1}{2} \log \left( 1 + \frac{\text{SNR}}{\frac{1}{3} \left[ Q^{-1} \left( \frac{p_e}{(1-2^{-R})} \right) \right]^2} \right), \quad (4)$$

which can be slightly relaxed to obtain a lower bound on  $R$  by

$$R > \frac{1}{2} \log \left( 1 + \frac{\text{SNR}}{\Gamma_0(p_e)} \right), \quad (5)$$

where

$$\Gamma_0(p_e) \stackrel{\text{def}}{=} \frac{1}{3} \left[ Q^{-1} \left( \frac{p_e}{2} \right) \right]^2. \quad (6)$$

Comparing (5) and (1), we see that PAM signaling with error probability  $p_e$  admits a capacity gap of at most  $\Gamma_0(p_e)$ . Comparing (4) and (5) it is clear that this upper bound is tight as  $R$  increases. For a typical value of  $p_e = 10^{-6}$ , the capacity gap of uncoded PAM is at most  $\Gamma_{0,\text{dB}} = 9\text{dB}$ .

Finally, we assume as usual that bits are mapped to PAM constellation symbols via Gray labeling. The associated bit error probability can thus be bounded by

$$p_b < \frac{2}{R} Q\left(\frac{\sqrt{P}\eta}{\sigma}\right) + 2Q\left(3\frac{\sqrt{P}\eta}{\sigma}\right) \approx \frac{p_e}{R}. \quad (7)$$

The bound follows by noting that erring toward the nearest neighbor incurs an error in a single bit, and by taking a worst case assumption for all other error events. The approximation becomes tight for small  $p_e$  due to the strong decay of the Q-function.

### B. Joint Source-Channel-Coding (JSCC) via Analog Transmission

It is well known [18] that when a Gaussian source is to be transmitted over an AWGN channel under a quadratic distortion measure, analog transmission obtains the optimal distortion (given by equating the rate-distortion function to

the channel capacity) with minimal delay. This solution is a simple instance of joint source-channel coding (JSCC). More explicitly, we wish to convey a Gaussian r.v.  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  over an AWGN channel  $Y = X + Z$ ,  $Z \sim \mathcal{N}(0, \sigma^2)$ , with expected input power constraint  $\mathbb{E}X^2 \leq P$  (i.e.  $\text{SNR} = \frac{P}{\sigma^2}$ ). The optimal transmission and estimation boil down to  $X = \alpha\varepsilon$  and  $\hat{\varepsilon} = \beta Y$ . The optimal choice of  $\alpha$  yields a power scaling factor, i.e.  $\alpha = \frac{\sqrt{P}}{\sigma_\varepsilon}$ , and optimal choice of  $\beta$  yields the Wiener coefficient  $\beta = \frac{\sigma_\varepsilon}{\sigma} \frac{\sqrt{\text{SNR}}}{\text{SNR}+1}$ . Plugging  $\alpha$  and  $\beta$  yields the minimal attainable MSE in this setup:

$$\mathbb{E}(\hat{\varepsilon} - \varepsilon)^2 = \frac{\sigma_\varepsilon^2}{\text{SNR} + 1} \quad (8)$$

Namely, this JSCC scheme improves the estimation error of  $\varepsilon$  by a factor  $\text{SNR} + 1$  relative to a trivial guess. In the sequel we shall use this simple construction as a building block for both the classic S-K scheme and the newly proposed noisy feedback schemes.

### C. The S-K Scheme

Consider the setting of communication over the AWGN channel with noiseless feedback, i.e., where  $\tilde{\sigma}^2 = 0$ . The S-K scheme can be described as follows. First, Terminal A maps the message  $W$  to the real-valued variable  $\Theta$  using a PAM modulation of size  $2^{NR}$ . In the first round, it sends a scaled version of  $\Theta$  satisfying the power constraint  $P$ . In subsequent rounds, Terminal B maintains an estimate  $\hat{\Theta}_n$  of  $\Theta$  given all the observation it has, and feeds it back to Terminal A. Terminal A then computes the estimation error  $\varepsilon_n \stackrel{\text{def}}{=} \hat{\Theta}_n - \Theta$ , and sends it to Terminal B using analog transmission.

(A) Initialization:

**Terminal A:** Map the message  $W$  to a PAM point  $\Theta$ .

**Terminal A  $\Rightarrow$  Terminal B:**

- Send  $X_1 = \sqrt{P}\Theta$
- Receive  $Y_1 = X_1 + Z_1$

**Terminal B:** Initialize the  $\Theta$  estimate<sup>1</sup> to  $\hat{\Theta}_1 = \frac{Y_1}{\sqrt{P}}$ .

(B) Iteration:

**Terminal B  $\Rightarrow$  Terminal A:**

- Send the current  $\Theta$  estimate:  $\tilde{X}_n = \hat{\Theta}_n$
- Receive  $\tilde{Y}_n = \tilde{X}_n$

**Terminal A:** Compute the estimation error  $\varepsilon_n = \tilde{Y}_n - \Theta$ .

**Terminal A  $\Rightarrow$  Terminal B:**

- Send  $\varepsilon_n$  via analog transmission. i.e.  $X_{n+1} = \alpha_n \varepsilon_n$ , where  $\alpha_n = \frac{\sqrt{P}}{\sigma_n}$  where  $\sigma_n^2 \stackrel{\text{def}}{=} \mathbb{E}\varepsilon_n^2$ .
- Receive  $Y_{n+1} = X_{n+1} + Z_{n+1}$

**Terminal B:** Update the  $\Theta$  estimate<sup>1</sup>  $\hat{\Theta}_{n+1} = \hat{\Theta}_n - \varepsilon_n$ , where

$$\hat{\varepsilon}_n = \beta_{n+1} Y_{n+1}$$

<sup>1</sup>Note that this is the minimum variance unbiased estimate of  $\Theta$ .

is the *Minimum Mean Squared Error* (MMSE) estimate of  $\varepsilon_n$ , thus,  $\beta_{n+1}$  is the appropriate Wiener coefficient:

$$\beta_{n+1} = \frac{\sqrt{P\sigma_n^2}}{P + \sigma^2} = \frac{\sigma_n}{\sigma} \cdot \frac{\sqrt{\text{SNR}}}{1 + \text{SNR}}.$$

(C) Decoding:

At time  $N$ , Terminal B decodes the message using a minimum distance decoder for  $\hat{\Theta}_N$  w.r.t. the PAM constellation.

To calculate the error probability and rate attained by the S-K scheme, we note that  $\varepsilon_{n+1} = \varepsilon_n - \hat{\varepsilon}_n$ . Using the property (8) of analog transmission yields:

$$\sigma_{n+1}^2 = \frac{\sigma_n^2}{1 + \text{SNR}} = \frac{1}{\text{SNR}(1 + \text{SNR})^n}. \quad (9)$$

An important observation is that using (9) and the fact that the power of  $\Theta$  is normalized to unity, one can regard the channel from  $\Theta$  to  $\Theta_N$  as an AWGN channel with a signal-to-noise ratio  $\text{SNR}_N = \sigma_N^{-2}$ , namely:

$$\text{SNR}_N = \text{SNR} \cdot (1 + \text{SNR})^{N-1}. \quad (10)$$

Note that it is possible to use the (biased) MMSE at the first round and compensate for the bias in the last round [19], yielding an end-to-end signal-to-noise ratio of  $(1 + \text{SNR})^N - 1$ . However, this improvement is negligible and also complicates the analysis in the sequel. Plugging  $\text{SNR}_N$  into (3) and bounding the Q-function by  $Q(x) < \frac{1}{2} \exp(-\frac{1}{2}x^2)$  gives:

$$p_e < \exp\left(-\frac{3}{2} \frac{\text{SNR} \cdot (1 + \text{SNR})^{N-1}}{2^{2NR} - 1}\right).$$

Plugging in the AWGN channel capacity (1) and removing the “-1” term, we obtain:

$$p_e < \exp\left(-\frac{3}{2} \frac{\text{SNR}}{1 + \text{SNR}} \cdot 2^{2N(C-R)}\right).$$

which is the well-known doubly exponential decay of the error probability of the S-K scheme.

Let us now provide an alternative interpretation of the S-K scheme performance, in terms of the capacity gap attained after a finite number of rounds. Plugging  $\text{SNR}_N$  in (5) yields:

$$R > \frac{1}{2N} \log\left(1 + \frac{\text{SNR} \cdot (1 + \text{SNR})^{N-1}}{\Gamma}\right).$$

Substituting the resulting  $R$  in the definition of the capacity gap (2) and assuming  $\text{SNR} \gg 1$  yields the following approximation for high SNR:

$$\Gamma_{\text{dB}}^{\text{S-K}}(p_e, N) \approx \frac{\Gamma_{0,\text{dB}}(p_e)}{N}.$$

This behavior is depicted by the dashed curve in Fig. 4.

## D. Joint Source-Channel Coding with Side Information

A key observation made in this work is that while the transmission of  $\varepsilon_n$  over the feedforward link (i.e. from Terminal A to Terminal B) can be regarded as a JSCC problem, the transmission of  $\hat{\Theta}_n$  over the feedback link (i.e. from Terminal B to Terminal A) can be regarded as a JSCC problem with side information. More explicitly, at round  $n$ , Terminal B holds its current estimate  $\hat{\Theta}_n = \Theta + \varepsilon_n$  and wants to convey it to Terminal A, while Terminal A knows  $\Theta$  and can use it as side information. To exploit this, we employ a lattice-based JSCC scheme with side information based on a more general scheme by Kochman and Zamir [9]. We note that for clarity of exposition and ease of analysis, we use the high-SNR version of [9], which can be slightly suboptimal in the low-SNR regime.

Let us now describe this solution. For simplicity, we start with the scalar case. First, we need some definitions and properties of modulo arithmetic. For a given  $d > 0$ , the scalar modulo- $d$  function is

$$\mathbb{M}_d[x] \stackrel{\text{def}}{=} x - d \cdot \text{round}\left(\frac{x}{d}\right)$$

where the  $\text{round}(\cdot)$  operator returns nearest integer to its argument (rounding up at half). The following properties are easily verified.

*Proposition 1:* The following properties hold:

- (i)  $\mathbb{M}_d[x] \in [-\frac{d}{2}, \frac{d}{2})$
- (ii) The *modulo distributive law*

$$\mathbb{M}_d[\mathbb{M}_d[x + d_1] + d_2 - x] = \mathbb{M}_d[d_1 + d_2]$$

- (iii) if  $d_1 + d_2 \in [-\frac{d}{2}, \frac{d}{2})$ , then

$$\mathbb{M}_d[\mathbb{M}_d[x + d_1] + d_2 - x] = d_1 + d_2. \quad (11)$$

otherwise, a *modulo-aliasing error* term of  $kd$  is added to the right-hand-side (11), for some integer  $k \neq 0$ .

- (iv) Let  $V \sim \text{Uniform}([-\frac{d}{2}, \frac{d}{2}))$ . Then  $\mathbb{M}_d[x + V]$  is uniformly distributed over  $[-\frac{d}{2}, \frac{d}{2})$  for any  $x \in \mathbb{R}$ .
- (v)  $\mathbb{E}(\mathbb{M}_d[x + V])^2 = \frac{d^2}{12}$ .

Using the above properties, we can provide the following solution to the JSCC with side information problem over the feedback link. Terminal B is in possession of  $\hat{\Theta} = \Theta + \varepsilon$  and wishes to convey it to Terminal A, where  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  and  $\Theta$  is known to Terminal A. The terminals are connected through an AWGN channel:  $\tilde{Y} = \tilde{X} + \tilde{Z}$ , with the appropriate noise distribution  $\tilde{Z} \sim \mathcal{N}(0, \tilde{\sigma}^2)$  and an input power constraint  $\mathbb{E}\tilde{X}^2 \leq \tilde{P}$ . Let  $V \sim \text{Uniform}([-\frac{d}{2}, \frac{d}{2}))$  be a *dither* signal known at both terminals. Then, Terminal B transmits

$$\tilde{X} = \mathbb{M}_d[\gamma\hat{\Theta} + V],$$

where we set  $d = \sqrt{12\tilde{P}}$  to guarantee that the power constraint is satisfied. Terminal A computes the estimate

$$\hat{\varepsilon} = \frac{1}{\gamma} \mathbb{M}_d [\tilde{Y} - \gamma\Theta - V].$$

Hence, by Proposition 1 property (ii):

$$\hat{\varepsilon} = \frac{1}{\gamma} \mathbb{M}_d [\gamma\varepsilon + \tilde{Z}]. \quad (12)$$

In the case where  $\gamma\varepsilon + \tilde{Z} \in [-\frac{d}{2}, \frac{d}{2}]$  (Proposition 1 property (iii)) we obtain

$$\hat{\varepsilon} = \varepsilon + \frac{1}{\gamma} \tilde{Z}. \quad (13)$$

The question that arises at this point is how to set  $\gamma$ . Clearly, a large  $\gamma$  would increase the probability of a modulo-aliasing error, but at the same time would reduce the additive estimation error in  $\hat{\varepsilon}$ . Denoting the modulo-aliasing error probability by  $p_{\text{mod}}$ . Proposition 1 property (iii) implies that

$$p_{\text{mod}} \stackrel{\text{def}}{=} \Pr \left( \gamma\varepsilon + \tilde{Z} \notin [-\frac{d}{2}, \frac{d}{2}] \right).$$

Recalling that  $d = \sqrt{12\tilde{P}}$ , and that  $\varepsilon$  and  $\tilde{Z}$  are jointly Gaussian with known variances, we obtain

$$p_{\text{mod}} = 2Q \left( \sqrt{\frac{3\tilde{P}}{\gamma^2\sigma_\varepsilon^2 + \tilde{\sigma}^2}} \right). \quad (14)$$

Let us now introduce the *looseness parameter*  $L$ , defined as:

$$L = \frac{\tilde{P}}{\gamma^2\sigma_\varepsilon^2 + \tilde{\sigma}^2}. \quad (15)$$

Using this definition, we can write  $p_{\text{mod}}$  as

$$p_{\text{mod}} = 2Q \left( \sqrt{3L} \right)$$

Observe that a larger  $L$  implies a smaller variance of the modulo argument in (12), hence a smaller modulo-aliasing error probability. On the other hand, a larger  $L$  implies a smaller  $\gamma$ , and hence a larger estimation error by virtue of (13). In the sequel, it will be convenient to express our results in terms of  $L$  instead of  $\gamma$ , as the former is a more natural parameter of the problem.

## V. SIMPLE INTERACTION

In this section we present our simplicity-oriented interactive scheme, using the S-K scheme and the scalar modulo JSCC scheme with side information as building blocks. We analyze the associated capacity gap and discuss implementation issues. The scheme is presented in Subsection V-A. An upper bound on the capacity gap attained by the scheme is given in Subsection V-B, and proved in Subsection V-C. Numerical results are presented in Subsection V-D. Practical implementation issues are addressed in Subsection V-E, and a concluding discussion appears in Subsection V-F.

### A. The Proposed Scheme

In what follows, we assume that the terminals share a common random i.i.d sequence  $\{V_n\}_{n=1}^N$ , mutually independent of the noise sequences and the message, where  $V_n \sim \text{Uniform}([-\frac{d}{2}, \frac{d}{2}])$ . As before, we set  $d = \sqrt{12\tilde{P}}$ . Recall the definition of  $\hat{\Theta}_n$  and  $\varepsilon_n$  in Subsection IV-C, as the estimator of  $\Theta$  and the corresponding estimation error at Terminal B at time  $n$ .

A block diagram of the scheme is depicted in Fig. 2. Let us describe our scheme in detail. The scheme is given in terms of the parameters  $\alpha, \beta_n, \gamma_n$  which dictate the performance. The specific choice of these parameters is given in the next subsection.

#### (A) Initialization:

**Terminal A:** Map the message  $W$  to a PAM point  $\Theta$ .

**Terminal A  $\Rightarrow$  Terminal B:**

- Send  $X_1 = \sqrt{P}\Theta$
- Receive  $Y_1 = X_1 + Z_1$

**Terminal B:** Initialize the  $\Theta$  estimate to  $\hat{\Theta}_1 = \frac{Y_1}{\sqrt{P}}$ .

#### (B) Iteration:

**Terminal B  $\Rightarrow$  Terminal A:**

- Given the  $\Theta$  estimate  $\hat{\Theta}_n$ , compute and send

$$\tilde{X}_n = \mathbb{M}_d [\gamma_n \hat{\Theta}_n + V_n]$$

- Receive  $\tilde{Y}_n = \tilde{X}_n + \tilde{Z}_n$

**Terminal A:** Extract a noisy version of estimation error  $\varepsilon_n$ :

$$\tilde{\varepsilon}_n = \frac{1}{\gamma_n} \mathbb{M}_d [\tilde{Y}_n - \gamma_n \Theta - V_n] \quad (16)$$

$$= \frac{1}{\gamma_n} \mathbb{M}_d [\gamma_n \varepsilon_n + \tilde{Z}_n] \quad (17)$$

Note that  $\tilde{\varepsilon}_n = \varepsilon_n + \frac{1}{\gamma_n} \tilde{Z}_n$ , unless a *modulo-aliasing error* occurs.

**Terminal A  $\Rightarrow$  Terminal B:**

- Send a scaled version of  $\tilde{\varepsilon}_n$ :  $X_{n+1} = \alpha \gamma_n \tilde{\varepsilon}_n$ , where  $\alpha$  is set to satisfy input power constraint  $P$  (computed later).
- Receive  $Y_{n+1} = X_{n+1} + Z_{n+1}$

**Terminal B:** Update the  $\Theta$  estimate  $\hat{\Theta}_{n+1} = \hat{\Theta}_n - \hat{\varepsilon}_n$ , where

$$\hat{\varepsilon}_n = \beta_{n+1} Y_{n+1}$$

The choice of  $\beta_n$  is described in the sequel.

#### (C) Decoding:

At time  $N$  the receiver decodes the message using a minimum distance decoder for  $\hat{\Theta}_N$  w.r.t. the PAM constellation.

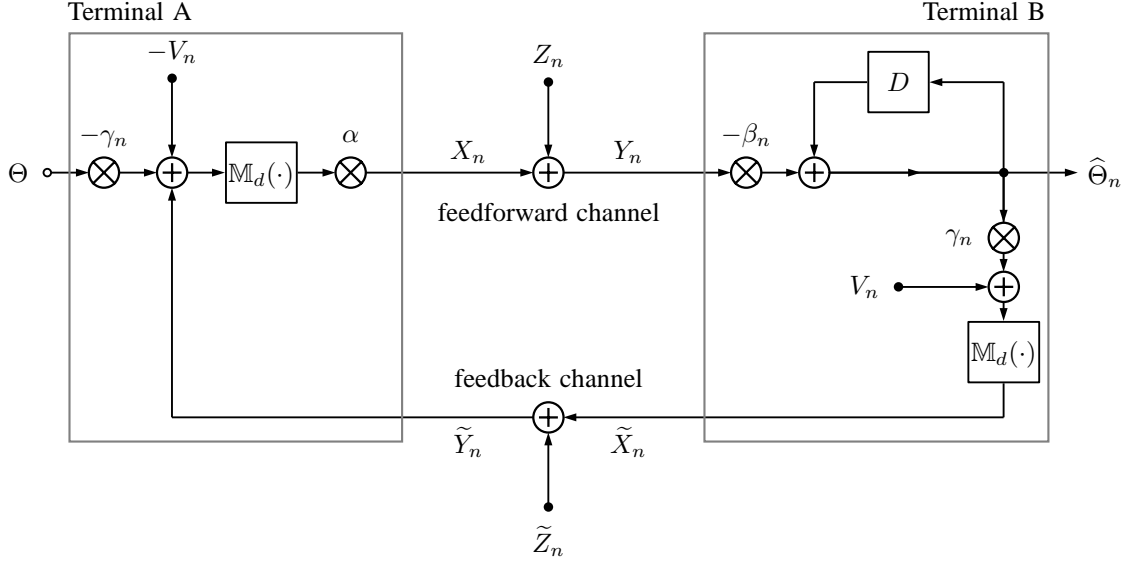


Fig. 2. Block diagram of the proposed scheme

### B. Main Result: The Capacity Gap

Recall the capacity gap function  $\Gamma_0(\cdot)$  of uncoded PAM given in (6). Fix a desired number of rounds  $N$ , a target error probability  $\delta$ , and set the target modulo-error probability to

$$p'_m = \frac{\delta}{2(N-1)}. \quad (18)$$

Set the looseness parameter to

$$L = \frac{1}{3} \left[ Q^{-1} \left( \frac{p'_m}{2} \right) \right]^2, \quad (19)$$

and set the scheme parameters  $\alpha, \beta_n, \gamma_n$  to

$$\alpha = \sqrt{\Psi_3^{-1} L \frac{P}{\tilde{P}}}, \quad (20)$$

$$\beta_n = \frac{\sigma_{n-1}}{\sigma} \frac{\sqrt{\Psi_3^{-1} \text{SNR} \cdot (1 - L \tilde{\text{SNR}}^{-1})}}{1 + \Psi_3^{-1} \text{SNR}},$$

$$\gamma_n = \sqrt{\frac{1}{\sigma_n^2} \left( \frac{\tilde{P}}{L} - \tilde{\sigma}^2 \right)},$$

where

$$\sigma_n^2 = \Psi_3 \text{SNR}^{-1} \left( 1 + \text{SNR} \cdot \Psi_3^{-1} \frac{1 - L \cdot \tilde{\text{SNR}}^{-1}}{1 + \Psi_3 L \cdot \Delta \text{SNR}^{-1}} \right)^{1-n}.$$

Define:

$$\Psi_1 \stackrel{\text{def}}{=} 1 + \Psi_3 L \cdot \Delta \text{SNR}^{-1} \quad (21)$$

$$\Psi_2 \stackrel{\text{def}}{=} \frac{1}{1 - L \cdot \tilde{\text{SNR}}^{-1}}$$

$$\Psi_3 \stackrel{\text{def}}{=} 1 + \frac{3}{2} L p'_m \left( N - 1 - \frac{2}{N} \right) \quad (22)$$

$$\Psi_4 \stackrel{\text{def}}{=} \exp \left( \frac{1}{\text{SNR} \cdot \Psi_3^{-1} (\Psi_1 \Psi_2)^{-\frac{N-1}{N}} \Gamma_0^{-\frac{1}{N}} \left( \frac{\delta}{2} \right) - 1} \right)$$

*Theorem 1:* For the choice of parameters above, the interactive communication scheme described in Subsection V-A achieves in  $N$  rounds an error probability  $p_e \leq \delta$  and a capacity gap  $\Gamma_{\text{dB}}^*$  satisfying:

$$\Gamma_{\text{dB}}^*(\delta, N) < \frac{1}{N} \Gamma_{0,\text{dB}} \left( \frac{\delta}{2} \right) + \frac{N-1}{N} (\Psi_{1,\text{dB}} + \Psi_{2,\text{dB}}) + \Psi_{3,\text{dB}} + \Psi_{4,\text{dB}}, \quad (23)$$

provided that  $\Psi_4 > 1$  (see Remark 5). The penalty term  $\Psi_3$  is typically negligible (see Remark 4).

We prove this theorem in Subsection V-C.

*Corollary 1 (High SNR behavior):* Let  $\Delta \text{SNR}$  and  $\delta$  be fixed. The capacity gap attained by our scheme for SNR large enough, can be approximated by

$$\Gamma_{\text{dB}}^*(\delta, N) \approx \frac{1}{N} \Gamma_{0,\text{dB}} \left( \frac{\delta}{2} \right) + \frac{N-1}{N} [1 + L \Delta \text{SNR}^{-1}]_{\text{dB}}. \quad (24)$$

Note that the first term is roughly the capacity gap of the S-K scheme with noiseless feedback, and that the second term depends only on  $\Delta \text{SNR}$ .

The following remarks are in order.

*Remark 2:* The penalty term  $\Psi_1$  can be attributed to noise insertion from the feedback channel to the feedforward channel. As shown below in (33), our feedforward transmission can

be interpreted as analog transmission with an additional noise term emanating from the feedback channel. The aggregate noise variance grows from  $\sigma^2$  to  $\Psi_1\sigma^2$ , causing a decrease in the signal-to-noise ratio, and a corresponding increase in the capacity gap.

*Remark 3:* The penalty term  $\Psi_2$  can be attributed to *power loss*, which is a consequence of the noise insertion discussed above. Recall that Terminal A sends  $\alpha\gamma_n\tilde{\varepsilon}_n$  which is a (scaled) noisy version of the true estimation error  $\varepsilon_n$ , and where the noise stems from the feedback channel. Thus, part of the transmission power  $P$  is consumed by this noise, leaving only a power  $P/\Psi_2 < P$  for the description of  $\varepsilon_n$ . This in turn reduces the signal-to-noise ratio and correspondingly increases the capacity gap.

*Remark 4:* The penalty term  $\Psi_3$  can be attributed to the deviation in the feedforward transmission power caused by modulo-aliasing errors. It is important to note that  $\Psi_{3,\text{dB}} = 0\text{dB}$  for  $N < 3$ . Moreover,  $\Psi_{3,\text{dB}}$  is typically negligible for  $N \geq 3$  and practical settings of  $\delta$ . This can be seen as follows: Recalling (19) and using the inverse form of the exponential bound for the  $Q$  function,  $Q^{-1}(x) \leq \sqrt{-2\ln(2x)}$ , yields

$$\begin{aligned}\Psi_3 &\leq 1 + \frac{3}{2}\sqrt{-2\ln\left(\frac{\delta}{N-1}\right)} \cdot \frac{\delta}{2(N-1)}\left(N-1-\frac{2}{N}\right) \\ &\leq 1 + \frac{3}{4}\delta\sqrt{-2\ln\left(\frac{\delta}{N-1}\right)}.\end{aligned}$$

Setting e.g.  $\delta = 10^{-6}$  and  $N = 10$  in the above equation yields  $\Psi_{3,\text{dB}} \leq 2 \cdot 10^{-5}\text{dB}$ , which is clearly negligible for all practical purposes.

*Remark 5:*  $\Psi_4$  is an additional penalty term, that results from the fact that we consider the capacity gap in terms of SNR ratios, whereas the explicit term arising from the capacity formula is related to  $\log(1 + \text{SNR})$  rather than  $\log(\text{SNR})$ . Note that  $\Psi_{4,\text{dB}} = O(\text{SNR}^{-1})$ . Due to simplifying assumptions taken in the bounding technique it is necessary to assume that  $\Psi_{4,\text{dB}} > 0$ . The case where  $\Psi_{4,\text{dB}} < 0$  corresponds to settings of very low SNR/ low  $\Delta\text{SNR}$ / very low  $\delta$  which is a less interesting regime. Note that in this case the capacity gap can still be calculated, but the bound is more cumbersome and is left out.

*Remark 6:* Note that there is a “low SNR” regime (related also to the target error probability or to  $\Delta\text{SNR}$ ), where the loss terms  $\Psi_{1,\text{dB}} + \Psi_{2,\text{dB}}$  are larger than say  $\Gamma_{0,\text{dB}}(\delta)$ . In that case, setting  $N = 1$ , namely using an uncoded system with no interaction, is the optimal choice of parameters for our scheme.

### C. Proof of Theorem 1

In Subsection IV-C we analyzed the error probability of the S-K scheme with noiseless feedback, relying on the fact that all the noises are jointly Gaussian, including the noise  $\varepsilon_N$  experienced by the PAM decoder. To that end, we were able to directly use the error probability analysis of simple PAM over an AWGN channel discussed in Subsection IV-A.

In the noisy feedback case however, the non-linearity of the modulo operations at both terminals induces a non-Gaussian distribution of  $\varepsilon_N$ . An analysis of the decoding error based on the actual distribution of  $\varepsilon_N$  seems involved. Yet, an upper bound can be derived via a simple coupling argument described below.

Recall that Terminal A computes  $\tilde{\varepsilon}_n$ , a noisy version of the estimation error at Terminal B, via a modulo operation (16). For any  $n \in \{1, \dots, N-1\}$  we define  $E_n$  as the event where this computation results in a modulo-aliasing error, i.e.,

$$E_n \stackrel{\text{def}}{=} \{\gamma_n\varepsilon_n + \tilde{Z}_n \notin [-\frac{d}{2}, \frac{d}{2}]\}. \quad (25)$$

Furthermore, we define  $E_N$  to be the PAM decoding error event:

$$E_N = \{\varepsilon_N \notin [-\frac{d_{\min}}{2}, \frac{d_{\min}}{2}]\},$$

where  $d_{\min}$  is the minimal distance of the PAM. The error probability of our scheme is clearly  $p_e \stackrel{\text{def}}{=} \Pr(E_N)$ . However, as mentioned above, the distribution of  $\varepsilon_N$  is not Gaussian due to the nonlinearity introduced by the modulo operations, which in turn renders  $p_e$  difficult to compute. To circumvent this problem, we instead consider the following trivial upper bound

$$p_e \leq \Pr\left(\bigcup_{n=1}^N E_n\right). \quad (26)$$

We intuitively expect this bound to be rather tight, since a modulo-aliasing error is very likely to imply a PAM decoding error. It turns out that upper bounding the right-hand-side above is not too difficult, as we now show.

To proceed, we define the *coupled system* as a system that is fed by the exact same message, and experiences the (sample-path) exact same noises. The coupled system differs from the original system in two aspects: first, it does not apply any modulo operations at either of the terminals; second, its feedforward power constraint is set to be  $P' = P/\Psi_3$  (where we appropriately define  $\text{SNR}' \stackrel{\text{def}}{=} \text{SNR}/\Psi_3$ ). We denote all the signals and events in the coupled system using the same notation as in the original system, but with an additional prime (') symbol, unless the signals are always identical between the systems by construction (e.g.  $Z_n$ ). Clearly, the coupled system can violate the power constraint at Terminal B. However, given the message  $W$ , all the random variables in the coupled system



are jointly Gaussian, and in particular, the estimation errors  $\varepsilon'_n$  in that system are Gaussian for  $n = 1, \dots, N$ . Moreover, it is easy to see that given no modulo-aliasing has occurred up to time  $n$ , the estimation errors are *sample-path identical* between the systems, i.e.  $\varepsilon'_n = \varepsilon_n$ . This leads to the following lemma:

*Lemma 1:* For any  $N > 1$ :

$$\Pr\left(\bigcup_{n=1}^N E_n\right) = \Pr\left(\bigcup_{n=1}^N E'_n\right).$$

*Proof:* Define the event

$$J_n \stackrel{\text{def}}{=} \bigcap_{i=1}^n \overline{E_i}$$

Let us show by induction that  $J_n = J'_n$ . For  $n = 1$ , we have

$$\begin{aligned} J_1 &= \{\gamma_1 \varepsilon_1 + \tilde{Z}_1 \in [-\frac{d}{2}, \frac{d}{2}]\} \\ &= \{\gamma_1 \varepsilon'_1 + \tilde{Z}_1 \in [-\frac{d}{2}, \frac{d}{2}]\} \\ &= J'_1 \end{aligned} \quad (27)$$

where (27) follows from the sample path identity. Assuming  $J_{k-1} = J'_{k-1}$  and using the sample path identity again, we have

$$\begin{aligned} J_k &= \{\gamma_k \varepsilon_k + \tilde{Z}_k \in [-\frac{d}{2}, \frac{d}{2}]\} \cap J_{k-1} \\ &= \{\gamma_k \varepsilon'_k + \tilde{Z}_k \in [-\frac{d}{2}, \frac{d}{2}]\} \cap J'_{k-1} \\ &= J'_k \end{aligned}$$

By the exact same argument (replacing  $\in$  with  $\notin$ ) we clearly have that  $J_{n-1} \cap E_n = J'_{n-1} \cap E'_n$ . Thus we can write

$$\begin{aligned} \Pr\left(\bigcup_{n=1}^N E_n\right) &= \Pr(E_1) + \sum_{n=2}^N \Pr\left(\bigcap_{i=1}^{n-1} \overline{E_i} \cap E_n\right) \\ &= \Pr(\overline{J_1}) + \sum_{n=2}^N \Pr(J_{n-1} \cap E_n) \\ &= \Pr(\overline{J_1}) + \sum_{n=2}^N \Pr(J'_{n-1} \cap E'_n) \\ &= \Pr\left(\bigcup_{n=1}^N E'_n\right) \end{aligned}$$

■

Combining the above with (26) and applying the union bound in the coupled system, we obtain

$$p_e \leq \sum_{n=1}^N \Pr(E'_n). \quad (28)$$

Thus, we can now upper bound the error probability by calculating probabilities in the coupled system, which involves only scalar Gaussian densities and significantly simplifies the analysis.

Let us begin by calculating the scheme parameters, and then use them to calculate the capacity gap. As mentioned

in Subsection V-B, the target error probability is set to  $\delta$ . Let us set  $\Pr(E'_N) = \frac{\delta}{2}$ , and  $\Pr(E'_1) = \dots = \Pr(E'_{n-1}) \stackrel{\text{def}}{=} p'_m$ , where  $p'_m$  is given in (18), and calculate the corresponding  $L$ . Recalling the definition of the event  $E'_n$  in (25) and that  $d = \sqrt{12\tilde{P}}$ , and since  $\gamma_n \varepsilon'_n + \tilde{Z}_n$  is Gaussian, we have that

$$p'_m = 2Q\left(\sqrt{\frac{3\tilde{P}}{\mathbb{E}(\gamma_n \varepsilon'_n + \tilde{Z}_n)^2}}\right).$$

Using the definition of  $L$  in (15), where here  $\sigma_\varepsilon^2 = \sigma_n^2 = \mathbb{E}(\varepsilon'_n)^2$  yields

$$\mathbb{E}(\gamma_n \varepsilon'_n + \tilde{Z}_n)^2 = \gamma_n^2 \sigma_n^2 + \tilde{\sigma}^2 = \frac{\tilde{P}}{L}. \quad (29)$$

First note that solving this equation for  $\gamma_n$  yields (20). Also note that combining the above two equations gives us the following setting of  $L$ :

$$L = \frac{1}{3} \left[ Q^{-1}\left(\frac{p'_m}{2}\right) \right]^2.$$

With  $L$  in hand, we can easily choose  $\alpha$  so that the power constraint at Terminal A is satisfied with equality. Namely

$$\begin{aligned} P' &= \mathbb{E}(X'_{n+1})^2 = \mathbb{E}(\alpha \gamma_n \tilde{\varepsilon}_n)^2 \\ &= \alpha^2 \mathbb{E}(\gamma_n \tilde{\varepsilon}_n)^2 = \alpha^2 \mathbb{E}(\gamma_n \varepsilon'_n + \tilde{Z}_n)^2. \end{aligned}$$

From (29) it follows that:

$$\alpha = \sqrt{L \frac{P'}{\tilde{P}}}. \quad (30)$$

The parameter  $\beta_n$  determines the evolution of the estimation error. The linear estimate of  $\varepsilon'_n$ :  $\hat{\varepsilon}'_n = \beta_{n+1} Y'_{n+1}$ , is the optimal estimate in the coupled system, in which  $\varepsilon'_n$  and  $Y'_{n+1}$  are jointly Gaussian. We would thus like to minimize  $\mathbb{E}(\varepsilon'_n - \hat{\varepsilon}'_n)^2$ . Recalling the input-output relation of the feed-forward channel and using (30) we obtain

$$Y'_{n+1} = \sqrt{\frac{LP'}{\tilde{P}}} (\gamma_n \varepsilon'_n + \tilde{Z}_n) + Z_{n+1} \quad (31)$$

and solving the optimization for  $\beta_n$  yields:

$$\beta_{n+1} = \frac{\sigma_n}{\sigma} \frac{\sqrt{\text{SNR}' \cdot (1 - L \text{SNR}^{-1})}}{1 + \text{SNR}'},$$

Noting that  $\varepsilon'_{n+1} = \varepsilon'_n - \hat{\varepsilon}'_n$  and computing the MMSE for the optimal choice of  $\beta_{n+1}$  above, we obtain a recursive formula for  $\sigma_n^2$  and  $\text{SNR}_n$ :

$$\begin{aligned} \text{SNR}_n &= \frac{1}{\sigma_n^2} \\ &= \text{SNR}' \cdot \left( 1 + \text{SNR}' \cdot \frac{1 - L \cdot \tilde{\text{SNR}}^{-1}}{1 + \Psi_3 L \cdot \Delta \text{SNR}^{-1}} \right)^{n-1} \\ &= \text{SNR}' \cdot (1 + \text{SNR}' \cdot \Psi_1 \Psi_2)^{n-1}. \end{aligned} \quad (32)$$

It is possible to give a different and modular interpretation for (32). Let us rewrite (31):

$$Y'_{n+1} = \sqrt{\frac{LP'}{P}} \gamma_n \varepsilon'_n + \sqrt{\frac{LP'}{P}} \tilde{Z}_n + Z_{n+1} \quad (33)$$

This equation can be regarded as a JSCC problem designated for the transmission of  $\varepsilon'_n$  over an AWGN channel. The effective noise of this AWGN channel is  $\sqrt{\frac{LP'}{P}} \tilde{Z}_n + Z_{n+1}$ . Some algebra shows that the variance of this noise is  $\Psi_1 \sigma^2$  where  $\Psi_1$  is defined in (21). We call this phenomenon *noise insertion*, as previously mentioned in Remark 2. A consequence of this phenomenon is that part of the transmission power is now consumed by the noise element related to  $\tilde{Z}_n$ . Subtracting this penalty from  $P'$  shows that the part of transmission power used for the description of  $\varepsilon'_n$  is  $P'/\Psi_2$ . We call this phenomena *power loss* as previously mentioned in Remark 3. So, all in all, the  $\text{SNR}'$  of the channel describe by (33) is  $\Psi_1 \Psi_2 \cdot \text{SNR}'$ . Using this fact together with the SNR evolution of the S-K scheme (10), and noting that the noise insertion and power loss effects only occur after the second round, we obtain (32).

We are now left with calculating the capacity gap. Let us find the condition that guarantees our scheme operates within the target error probability requirement. Rewriting (28) we get

$$\begin{aligned} p_e &\leq \sum_{n=1}^N \Pr(E'_n) \\ &\leq (N-1)p'_m + \Pr(E'_N) \\ &\leq \frac{\delta}{2} + \Pr(E'_N) \\ &\leq \frac{\delta}{2} + 2Q\left(\sqrt{\frac{3\text{SNR}_N}{2^{2NR}-1}}\right), \end{aligned} \quad (34)$$

hence by setting

$$Q\left(\sqrt{\frac{3\text{SNR}_N}{2^{2NR}-1}}\right) = \frac{\delta}{4} \quad (35)$$

we obtain that  $p_e \leq \delta$  as desired. We are now in a position to derive a lower bound on the capacity gap attained by our scheme: we can rearrange (35) to obtain a lower bound on  $R$ , use the expression (32) for  $\text{SNR}_N$ , and plug this into the definition of the capacity gap (2). This yields (23), where  $\Psi_{4,\text{dB}}$  is a remainder term obtained by pedestrian manipulations using the inequality  $-\ln(1-x) \leq \frac{x}{1-x}$  for  $x \in (0,1)$ . Note that the result was obtained for the specific choice  $p'_m = \frac{\delta}{2(N-1)}$  of the modulo-aliasing error. In general, reducing  $p_m$  increases  $L$  which in turn decreases  $\text{SNR}_N$ , and hence increases the second addend on the right-hand-side of (34), resulting in a trade-off that could potentially be further optimized.

We are now left with specifying the relation between  $P'$  and  $P$ , captured by  $\Psi_3$ . We start by defining the following sequence of events:

$$A_n \stackrel{\text{def}}{=} \{X_n = X'_n\} \quad \text{for } n = 1, \dots, N.$$

Using this definition, we can write

$$\begin{aligned} \mathbb{E}(X_n^2) &= \Pr(A_n) \mathbb{E}(X_n^2 | A_n) + \Pr(\overline{A_n}) \mathbb{E}(X_n^2 | \overline{A_n}) \\ \mathbb{E}(X'_n)^2 &= \Pr(A_n) \mathbb{E}((X'_n)^2 | A_n) + \Pr(\overline{A_n}) \mathbb{E}((X'_n)^2 | \overline{A_n}). \end{aligned}$$

Combining these expressions, while noting that  $\mathbb{E}(X'_n)^2 = P'$  and that  $\mathbb{E}((X'_n)^2 | A_n) = \mathbb{E}(X_n^2 | A_n)$ , we obtain:

$$\begin{aligned} \mathbb{E}(X_n^2) &= P' + \Pr(\overline{A_n}) \cdot \mathbb{E}(X_n^2 - (X'_n)^2 | \overline{A_n}) \\ &\leq P' + \Pr(\overline{A_n}) \cdot \mathbb{E}(X_n^2 | \overline{A_n}) \end{aligned} \quad (36)$$

Let us now calculate (36) separately for  $n < 3$  and  $n \geq 3$ . For  $n = 1$  we have  $X_1 = X'_1$  by construction, hence  $\Pr(\overline{A_1}) = 0$  and thus  $\mathbb{E}(X_1)^2 = P'$ . For  $n = 2$  we have by (17):

$$\begin{aligned} X_n &= \alpha \mathbb{M}_d [\gamma_1 \varepsilon_1 + \tilde{Z}_1] \\ X'_n &= \alpha (\gamma_1 \varepsilon_1 + \tilde{Z}_1). \end{aligned}$$

By definition of the scalar modulo operation it holds that  $|\mathbb{M}_d[x]| \leq |x|$ , and therefore  $|X_n| \leq |X'_n|$ , which implies that  $\mathbb{E}(X_n)^2 \leq P'$ . We are now left with the more general case of  $n \geq 3$ . We have that

$$\begin{aligned} \mathbb{E}(X_n^2 | \overline{A_n}) &\leq \mathbb{E}\left(\left(\alpha \mathbb{M}_d [\gamma_n \varepsilon_n + \tilde{Z}_n]\right)^2 | \overline{A_n}\right) \\ &= \left(\alpha \frac{d}{2}\right)^2 \\ &= \left(\sqrt{L \frac{P'}{\tilde{P}}} \frac{\sqrt{12\tilde{P}}}{2}\right)^2 \\ &= 3LP' \end{aligned}$$

Plugging the above in (36) yields

$$\mathbb{E}(X_n)^2 \leq P' (1 + 3 \Pr(\overline{A_n})L)$$

We are left with bounding  $\Pr(\overline{A_n})$  for  $n \geq 3$ . To that end, note that  $X_n \neq X'_n$  implies that at least one modulo error occurred up to time instant  $n-1$ , with probability one. Therefore:

$$\Pr(\overline{A_n}) \leq \Pr\left(\bigcup_{i=1}^{n-1} E_i\right) \leq \sum_{i=1}^{n-1} \Pr(E_i) = (n-1)p'_m$$

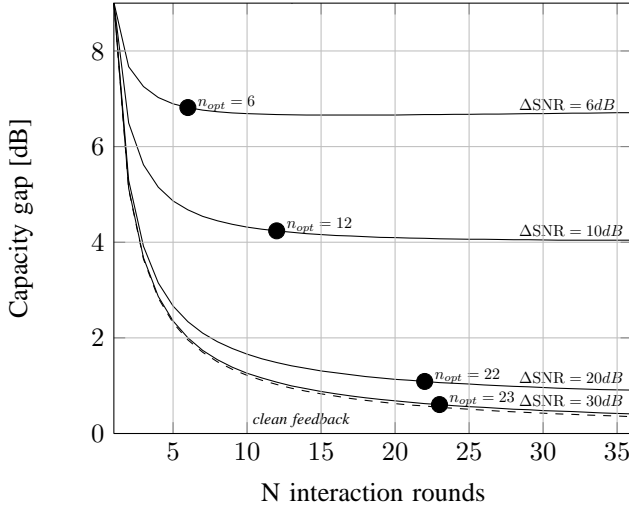


Fig. 3. The capacity gap as function of the iterations and  $\Delta\text{SNR}$  for a target rate  $R = 1$  (low SNR), and target error probability  $p_t = 10^{-6}$

Collecting the upper bounds on  $\mathbb{E}(X_n^2)$  for all  $n = 1, \dots, N$ , we obtain the following upper bound on the average power consumed by our scheme:

$$\begin{aligned}
 P &= \frac{1}{N} \sum_{n=1}^N \mathbb{E}(X_n^2) \\
 &\leq \frac{P'}{N} \left( 2 + \sum_{n=3}^N (1 + 3L(n-1)p'_m) \right) \\
 &= \frac{P'}{N} \left( N + \frac{3(N-2)(N+1)}{2} Lp'_m \right) \\
 &= P' \left( 1 + \frac{3}{2} Lp'_m \left( N - 1 - \frac{2}{N} \right) \right).
 \end{aligned}$$

Hence, setting  $P' = P/\Psi_3$  where  $\Psi_3$  is given by (22) obeys the feedforward transmission power constraint. This concludes the proof.

#### D. Numerical Results

The behavior of the capacity gap for our scheme as a function of the number of interaction rounds and  $\Delta\text{SNR}$  is depicted in Fig. 3 and Fig. 4, for high SNR and low SNR setups. In both figures we plotted the capacity gap, for a target rate  $R$  and a target error probability  $p_e = 10^{-6}$ , where the SNR is found by numerically solving  $R = \frac{1}{2N} \log \left( 1 + \frac{\text{SNR}_N}{\Gamma_0(\delta/2)} \right)$  and the capacity gap is calculated by definition. It can be seen that the higher the  $\Delta\text{SNR}$ , the smaller the capacity gap, where at  $\Delta\text{SNR} = 30\text{dB}$  we virtually obtain the noiseless feedback performance. The points marked  $n_{\text{opt}}$  are those for which the capacity gap is less than 0.2dB above the minimal value attained. In Fig. 3 the rate was set to  $R = 1$ , and it can be seen that  $\Delta\text{SNR} = 10\text{dB}$  reduces the capacity gap to 4.2dB in 12 iterations, and  $\Delta\text{SNR} = 20\text{dB}$  reduces the capacity gap to 1.1dB in 22 iterations. In Fig. 4, the rate was set to  $R = 4$ ,

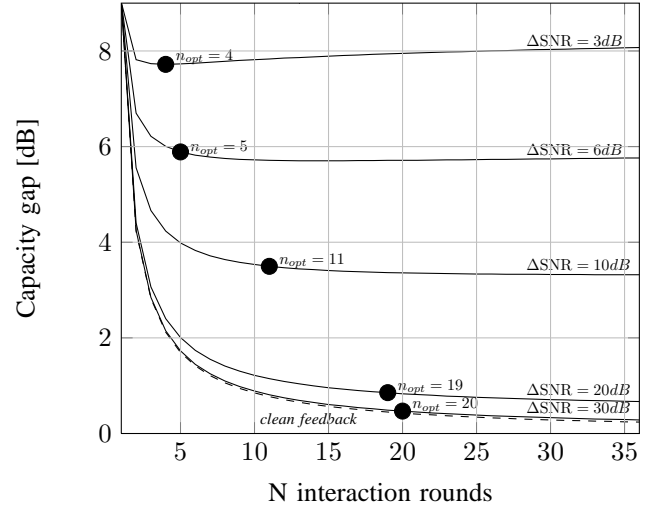


Fig. 4. The capacity gap as function of the iterations and  $\Delta\text{SNR}$  for a target rate  $R \geq 4$  (high SNR), and target error probability  $p_t = 10^{-6}$

and it can be seen that  $\Delta\text{SNR} = 10\text{dB}$  reduces the capacity gap to 3.5dB in 11 iterations, and  $\Delta\text{SNR} = 20\text{dB}$  reduces the capacity gap to 0.8dB in 19 iterations. Observing (24) we can see that for high SNR the result is only a function of  $\Delta\text{SNR}$ , thus does not depend on the target rate or the base SNR.

#### E. Notes on Implementation

The scheme described in this section is both simple and practical, as opposed to its noiseless feedback counterparts that break down in the presence of feedback noise. This provides impetus for further discussing implementation related aspects. The following conditions should be met for our results to carry merit:

- 1) *Information asymmetry*: Terminal A has substantially more information to convey than Terminal B.
- 2) *SNR asymmetry*: The SNR of the feedforward channel is lower than the SNR of the feedback channel. This can happen due to differences in power constraints (e.g. when Terminal A is battery operated and Terminal B is connected to the power grid), path losses, or noise/interference asymmetry.
- 3) *Complexity/delay constraints*: There are severe complexity or delay constraint at Terminal A.
- 4) *Two-way signaling*: Our scheme assumes sample-wise feedback. The communication system should therefore be full duplex where both terminals have virtually the same signaling rate; hence, the terminals split the bandwidth between them even though only Terminal A is transmitting information. This situation can sometimes be inherent to the system, but should otherwise be tested against the (non-interactive) solution where the entire bandwidth is allocated to Terminal A. This choice of forward vs. feedback

bandwidth allocation yields a system trade-off that is SNR dependent: Terminal A can use our scheme and achieve a rate of  $C(\text{SNR}_{dB} - \Gamma_{dB}^*)$  (where  $C(\cdot)$  is the Shannon capacity function), or alternatively employ non-interactive codes over the full forward-feedback bandwidth. The latter option doubles the forward signaling rate but also incurs a 3dB loss in SNR and a potentially larger capacity gap  $\Gamma_{dB}^\dagger$ , resulting in a rate of  $2C(\text{SNR}_{dB} - 3\text{dB} - \Gamma_{dB}^\dagger)$ . It can therefore be seen that our solution is generally better for low enough SNR. For instance, for  $\delta = 10^{-6}$  and  $\Delta\text{SNR} > 30\text{dB}$  our scheme outperforms (with comparable complexity and delay) full bandwidth uncoded PAM for any  $\text{SNR} < 23\text{dB}$ , and outperforms (with significantly smaller complexity and delay) full bandwidth non-feedback codes with  $\Gamma_{dB}^\dagger = 3\text{dB}$  for any  $\text{SNR} < 9\text{dB}$ .

- 5) *Bandwidth splitting*: We have tacitly assumed that the bandwidth is equally split between the feedforward and feedback channels. We note that a non-equal splitting of the bandwidth can also be handled. Standard techniques for trading bandwidth with SNR can be used in order to facilitate the use of our scheme in these scenarios.

The use of very large PAM constellations, whose size is exponential in the product of rate and interaction rounds, seemingly requires extremely low noise and distortion at the digital and analog circuits in Terminal A. This may appear to impose a major implementation obstacle. Fortunately, this is not the case. The full resolution implied by the constellation size is by construction confined *only* to the original message  $\Theta$  and the final estimate  $\hat{\Theta}_N$ ; the transmitted and received signals in the course of interaction can be safely quantized at a resolution determined only by the channel noise (and *not* the final estimation noise), as in commonplace communication systems. Figuratively speaking, the source bits are gradually *revealed* along the interaction process, where the number of bits revealed in every round is determined by the channel SNR.

Another important implementation issue is sensitivity to model assumptions. We have successfully verified in simulation the robustness of the proposed scheme in several reasonable scenarios including correlative noise, excess quantization noise, and multiplicative channel estimation noise. The universality of the scheme and its performance for a wider range of models remains to be further investigated.

#### F. Discussion

Note that so far we have limited our discussion to the PAM symbol. The bit-error rate is in fact lower, since an error in PAM decoding affects only a single bit with high probability (7), assuming Gray labeling. However, note that the modulo-aliasing error will typically result in many erroneous bits, and hence optimizing the bit error rate does not yield a major

improvement over its upper bound. Further fine-tuning of the scheme can be obtained by non-uniform power allocation over interaction rounds in both Terminal A and B; in particular, we note that Terminal B is silent in the last round, which can be trivially exploited.

We note again that for any choice of SNR and  $\Delta\text{SNR}$ , the error probability attained by our scheme cannot be made to vanish with the number interaction rounds while maintaining a non-zero rate, as in the noiseless feedback S-K scheme case. The reason is that (15) implies that  $L < \tilde{\text{SNR}}$ , which in turn by (14) imposes a lower bound on the attainable error probability, dictated by the probability of modulo-aliasing of the feedback noise. Equivalently, one cannot get arbitrarily close to capacity for a given target error probability, since increasing the number of iterations improves  $\text{SNR}_N$  and reduces the PAM decoding error term, but at the same time increases the modulo-aliasing error term in (34). Hence, our scheme is not *capacity achieving* in the usual sense. However, it can get close to capacity in the sense of reducing the capacity gap using a very short block length, typically  $N \approx 20$  in the examples presented. To the best of our knowledge, state-of-the-art (non-interactive) block codes require a block length typically larger by two orders of magnitude to reach the same gap at the same error probability. Consequently, the encoding delay of our scheme is markedly lower than that of these competing schemes. Alternatively, compared to a minimal delay uncoded system under the same error probability, our scheme operates at a much lower capacity gap for a wide regime of settings, and hence can be significantly more power efficient.

Another important issue is that of encoding and decoding complexity. Our proposed scheme applies only two multiplications and one modulo operation at each terminal in each interaction round. This is significantly lower than the encoding/decoding complexity of good block codes, even if non-optimal methods such as iterative decoding are employed.

## VI. IMPROVING RELIABILITY

In this section we describe two asymptotic versions of our simple scheme that are aimed at improving reliability at the cost of increased complexity and delay. These schemes are shown to outperform their non-feedback counterparts in the error exponent sense, when  $\Delta\text{SNR}$  is large enough. The first scheme is based on replacing the scalar PAM signaling with a random block code, and replacing the scalar modulo operation with a modulo-lattice operation. The error analysis of both the modulo error and the decoder decision error is performed in the coupled system as before, but are now concerned with error exponents instead of the scalar  $Q$ -function. The second scheme is based on concatenated coding

with the simple interaction scheme as an inner code and a random block outer code. Classical error exponent results are given in Subsection VI-A. JSCC with lattices is discussed in Subsection VI-B. The first scheme is introduced in Subsection VI-C. A lower bound on its error exponent is given in Subsection VI-D, and proved in Subsection VI-E. The second scheme is introduced in Subsection VI-F. A concluding discussion appears in Subsection VI-G.

#### A. Block Codes and Error Exponents

In what follows, we replace the scalar PAM mapping of the message point  $W \rightarrow \Theta$  with an AWGN channel block code mapping  $W \rightarrow \Theta$  of length  $N$ . In this subsection we cite the classic results on the performance of block-codes for the AWGN channel. For channel coding over the AWGN channel with a signal-to-noise ratio SNR and rate  $R$ , there exist block codes of length  $N$  whose average error probability (averaged over the messages) under maximum likelihood decoding is exponentially upper bounded by  $\Pr(\widehat{W}(\mathbf{Y}) \neq W) \leq e^{-NE_r(\text{SNR}, R)}$  where  $E_r(\text{SNR}, R)$  is given by [6]:

$$E_r(\text{SNR}, R) = \begin{cases} E_{sp}(\text{SNR}, R) & \text{if } R_{rc} < R \leq C \\ E_{rc}(\text{SNR}, R) & \text{if } R_{ex} < R \leq R_{rc} \\ E_{ex}(\text{SNR}, R) & \text{if } 0 < R \leq R_{ex} \end{cases}$$

The boundaries between the regions are as follows. The *Shannon capacity* is  $C \stackrel{\text{def}}{=} \frac{1}{2} \log(1 + \text{SNR})$ . The *critical rate* is  $R_{cr} \stackrel{\text{def}}{=} \frac{1}{2} \log \left( \frac{1}{2} + \frac{\text{SNR}}{4} + \frac{1}{2} \sqrt{1 + \frac{\text{SNR}^2}{4}} \right)$ . The *expurgation rate* is  $R_{ex} \stackrel{\text{def}}{=} \frac{1}{2} \log \left( \frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{\text{SNR}^2}{4}} \right)$ . The exponents in the above three regions are given by:

$$E_{sp}(\text{SNR}, R) = \frac{\text{SNR}}{4\beta} \left( \beta + 1 - (\beta - 1) \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} \right) + \frac{1}{2} \ln \left( \beta - \frac{\text{SNR}(\beta - 1)}{2} \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} \right) \quad (37)$$

where  $\beta = 2^{2R}$ ,

$$E_{rc}(\text{SNR}, R) = 1 - \beta + \frac{\text{SNR}}{2} + \frac{1}{2} \log \left( \beta - \frac{\text{SNR}}{2} \right) - \frac{1}{2} \log(\beta) - \log(2)R$$

where here  $\beta = 2e^{2R_{cr}}$ , and

$$E_{ex}(\text{SNR}, R) = \frac{\text{SNR}}{4} \left[ 1 - \sqrt{1 - 2^{-2R}} \right].$$

It is also well known and readily verified that for  $0 < R < C$ , the exponent  $E_{sp}(\text{SNR}, R)$  coincides with the asymptotic expression of Shannon's *sphere packing bound* for the AWGN channel [20]. Hence,  $E_{sp}(\text{SNR}, R)$  is also an upper bound for the reliability function, and is tight above the critical rate.

#### B. JSCC with Side Information Using High Dimensional Lattices

From this point on, boldface letters such as  $\mathbf{X}, \Theta$  denote vectors of size  $N_\Lambda$ . As shown in Subsection IV-D, the probability of modulo-aliasing error in the JSCC problem with side information in its scalar version, while sometimes small, is bounded away from zero. In order to make this probability arbitrarily small, the dimension of the scheme should be increased. This can be achieved by introducing large dimensional lattices and replacing the scalar modulo with the corresponding modulo-lattice operations. Let us start by quickly surveying a few basic lattice notation and properties [21]:

- (i) A lattice of dimension  $N_\Lambda$  is denoted  $\Lambda = G \cdot \mathbb{Z}^{N_\Lambda}$ , where  $G$  is the generating matrix.
- (ii)  $\text{Vol}(\Lambda) = |\det(G)|$  is the lattice cell volume.
- (iii) The nearest neighbor quantization of  $\mathbf{x}$  w.r.t.  $\Lambda$  is denoted  $\mathbb{Q}_\Lambda[\mathbf{x}]$ .
- (iv)  $\mathcal{V}_0 = \{\mathbf{x} : \mathbb{Q}_\Lambda[\mathbf{x}] = \mathbf{0}\}$  is the fundamental Voronoi cell pertaining to  $\Lambda$ .
- (v) The modulo  $\Lambda$  operation is  $\mathbb{M}_\Lambda[\mathbf{x}] \stackrel{\text{def}}{=} \mathbf{x} - \mathbb{Q}_\Lambda[\mathbf{x}]$ .
- (vi)  $\mathbb{M}_\Lambda[\cdot]$  satisfies the *distributive law*

$$\mathbb{M}_\Lambda[\mathbb{M}_\Lambda[\mathbf{x}] + \mathbf{y}] = \mathbb{M}_\Lambda[\mathbf{x} + \mathbf{y}]$$

- (vii) The volume to noise ratio (VNR) of a lattice in the presence of AWGN with variance  $\sigma^2$  is  $\mu(\Lambda) \stackrel{\text{def}}{=} [\text{Vol}(\Lambda)]^{2/N_\Lambda} / \sigma^2$ .
- (viii) The normalized second moment of a lattice  $\Lambda$  is  $G(\Lambda) \stackrel{\text{def}}{=} \sigma^2(\Lambda) / [\text{Vol}(\Lambda)]^{2/N_\Lambda}$ , where  $\sigma^2(\Lambda) = \frac{1}{N_\Lambda} \mathbb{E}(\|\mathbf{V}\|^2)$  and  $\mathbf{V}$  is uniformly distributed on  $\mathcal{V}_0$ .

Consider again the JSCC problem introduced in Subsection IV-D, where now Terminal B is in possession of a vector  $\hat{\Theta} = \Theta + \epsilon$ , and wants to convey the i.i.d.  $\sim \mathcal{N}(0, \sigma_\epsilon^2)$  error vector  $\epsilon$  to Terminal A which is in possession of  $\Theta$ , over an AWGN channel  $\tilde{Y} = \tilde{X} + \tilde{Z}$ . The channel is again characterized by  $\tilde{Z} \sim \mathcal{N}(0, \tilde{\sigma}^2)$  and  $\mathbb{E}\tilde{X}^2 \leq \tilde{P}$ . We assume that a *dither* signal  $\mathbf{V} \sim \text{Uniform}(\mathcal{V}_0)$ , mutually independent of the message and the channel noises, is known at both terminals.

Let us revise the JSCC with side information scheme presented in Subsection IV-D, replacing the scalar modulo operation  $\mathbb{M}_d[\cdot]$  with the lattice modulo operation  $\mathbb{M}_\Lambda[\cdot]$ . Hence, Terminal B transmits

$$\tilde{X} = \mathbb{M}_\Lambda[\gamma \hat{\Theta} + \mathbf{V}],$$

Terminal A estimates:

$$\hat{\epsilon} = \frac{1}{\gamma} \mathbb{M}_\Lambda[\tilde{Y} - \gamma \Theta - \mathbf{V}] = \frac{1}{\gamma} \mathbb{M}_\Lambda[\gamma \epsilon + \tilde{Z}]$$

And if  $\gamma\epsilon + \tilde{\mathbf{Z}} \in \mathcal{V}_0$  then

$$\hat{\epsilon} = \epsilon + \frac{1}{\gamma} \tilde{\mathbf{Z}}$$

We are now ready to set the parameters of the modulo-lattice scheme. Let us set the lattice's second moment to equal the feedback power constraint  $\sigma^2(\Lambda) = \tilde{P}$ . This would guarantee (due to dithering) that the feedback transmission power constraint is satisfied. The modulo-aliasing error event is the event where

$$\gamma\epsilon + \tilde{\mathbf{Z}} \notin \mathcal{V}_0$$

Recall the definition of the *looseness* parameter  $L$  of the lattice in (15), and note that  $L = \mu(\Lambda) \cdot G(\Lambda)$ . It was shown in [22, Theorem 5] that there exist lattices that asymptotically attain  $G(\Lambda) = \frac{1}{2\pi e} + o(1)$ , and a modulo-error that is exponentially bounded by  $p_{\text{mod}} \leq e^{-N_\Lambda E_p(\frac{\mu(\Lambda)}{2\pi e})}$ , where  $E_p(x)$  is the *Polytrev exponent*, given by [21]

$$E_p(x) = \begin{cases} \frac{1}{2}(x - 1 - \ln(x)) & \text{if } 1 < x \leq 2 \\ \frac{1}{2}(\ln(x) + \ln(\frac{e}{4})) & \text{if } 2 < x \leq 4 \\ \frac{1}{8}x & \text{if } x > 4 \end{cases}$$

Hence in our notation, such lattices satisfy

$$p_{\text{mod}} \leq e^{-N_\Lambda E_p(L)}. \quad (38)$$

Clearly, in a similar fashion to the scalar scheme, the setting of  $L$  determines the trade-off between modulo-error probability and decision error probability. Setting  $L$  close to 1 will maximize the effective signal-to-noise ratio (and maximize the block code error exponent), but at the same time minimize the modulo-error exponent. Setting  $L$  to be large will do the opposite, and will also reduce the maximal achievable rate. We note that the corresponding lattice-based JSCC scheme in [9] is better at low SNR, due to the addition of another Wiener coefficient multiplier at the receiver before the modulo operation, which results in non-Gaussian statistics of the error. For simplicity of analysis, this technique is not used here.

### C. Description of the Scheme

In this subsection, we show how to combine the blockwise coding and blockwise modulo operations into one scheme. In a nutshell, the message  $W$  is mapped into a codeword  $\Theta$  of length  $N_\Lambda$ , and sent in the first block ( $N_\Lambda$  channel uses). This replaces the PAM transmission in the scalar scheme. In the sequel, vector analog transmission is used over the feedforward, and vector modulo-lattice transmission is used over the feedback. Ultimately,  $W$  is decoded using a maximum likelihood decoding rule. Since under this protocol both terminals are idle half the time, we interlace two identical schemes, encoding and decoding two independent messages,

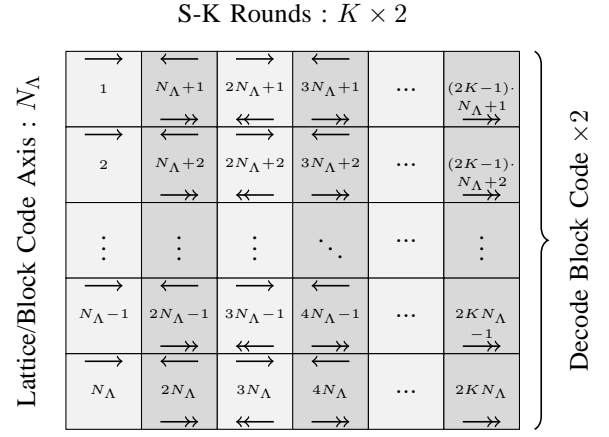


Fig. 5. Blockwise transmission. The time instants are divided into blocks of size  $N_\Lambda$ . Single headed arrows “→” and “←” denote transmission from Terminal A to Terminal B respectively. Double headed errors, “⇔” and “⇐”, bear the same meaning but for the second scheme.

as illustrated in Fig. 5. We denote the block index (or round index) by  $k \in \{1, \dots, K\}$ . For brevity, and with a mild abuse of notation, we only describe the evolution of one of the interlaced schemes.

The setting of the parameters  $\alpha, \beta_k, \gamma_k$  will be discussed in the sequel. The dither variables  $\mathbf{V}_k$  are i.i.d., uniformly distributed on  $\mathcal{V}_0$ , and mutually independent of the message and the noise processes.

(A) Initialization:

**Terminal A:** Map the message  $W$  to codeword  $\Theta$  using a codebook for the AWGN channel with average power  $P$ .

**Terminal A  $\Rightarrow$  Terminal B:**

- Send  $\mathbf{X}_1 = \Theta$
- Receive  $\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{Z}_1$

**Terminal B:** Initialize the  $\Theta$  estimate to  $\hat{\Theta}_1 = \mathbf{Y}_1$ .

(B) Iteration:

**Terminal B  $\Rightarrow$  Terminal A:**

- Given the  $\Theta$  estimate  $\hat{\Theta}_k$ , compute and send in the following block

$$\tilde{\mathbf{X}}_k = \mathbb{M}_\Lambda \left[ \gamma_k \hat{\Theta}_k + \mathbf{V}_k \right]$$

- Receive  $\tilde{\mathbf{Y}}_k = \tilde{\mathbf{X}}_k + \tilde{\mathbf{Z}}_k$

**Terminal A:** Extract a noisy scaled version of the estimation error vector  $\epsilon_k$ :

$$\tilde{\epsilon}_k = \frac{1}{\gamma_k} \mathbb{M}_\Lambda \left[ \tilde{\mathbf{Y}}_k - \gamma_k \Theta - \mathbf{V}_k \right]$$

Note that  $\tilde{\epsilon}_k = \epsilon_k + \frac{1}{\gamma_k} \tilde{\mathbf{Z}}_k$ , unless a modulo-aliasing error occurs.

**Terminal A  $\Rightarrow$  Terminal B:**

- Send a scaled version of  $\tilde{\epsilon}_k$ :  $\mathbf{X}_{k+1} = \alpha \gamma_k \tilde{\epsilon}_k$ , where  $\alpha$  is set so that the input power constraint  $P$  is met.

- Receive  $\mathbf{Y}_{k+1} = \mathbf{X}_{k+1} + \mathbf{Z}_{k+1}$

**Terminal B:** Update the  $\Theta$  estimate  $\hat{\Theta}_{k+1} = \hat{\Theta}_k - \hat{\epsilon}_k$ , where

$$\hat{\epsilon}_k = \beta_{k+1} \mathbf{Y}_{k+1}$$

- (C) Decoding: After the reception of block  $K$  the receiver decodes the message  $\hat{W}(\hat{\Theta}_K)$  using an ML decision rule w.r.t. the codebook.

#### D. Main Result: The Error Exponent

Set the scheme parameters  $\alpha, \beta_k, \gamma_k$  to

$$\alpha = \sqrt{L \frac{P}{\tilde{P}}},$$

$$\beta_k = \frac{\sigma_{k-1}}{\sigma} \frac{\sqrt{\text{SNR} \cdot (1 - L \tilde{\text{SNR}}^{-1})}}{1 + \text{SNR}},$$

$$\gamma_k = \frac{1}{\sigma_k} \sqrt{\frac{\tilde{P}}{L} - \tilde{\sigma}^2},$$

where  $\sigma_k = \sigma_k(L) = \frac{1}{\text{SNR}_K(L)}$ , and

$$\text{SNR}_K(L) \stackrel{\text{def}}{=} \text{SNR} \left( 1 + \text{SNR} \cdot \frac{1 - L \tilde{\text{SNR}}^{-1}}{1 + L \tilde{\Delta} \text{SNR}^{-1}} \right)^{K-1}, \quad (39)$$

The following theorem provides a lower bound on the error exponent obtained by our scheme.

*Theorem 2:* For the choice of parameters above, the interactive communication scheme described in Subsection VI-C with a total delay of  $N$  time instants, attains an error probability  $p_e \leq e^{-N E_{\text{FB}}(R)}$ , where

$$E_{\text{FB}}(R) \stackrel{\text{def}}{=} \max_{K \in \mathbb{N}, L \geq 1} \left\{ \frac{\min \{E_r(\text{SNR}_K(L), KR), E_p(L)\}}{2K} \right\}. \quad (40)$$

#### E. Proof of Theorem 2

Define the error event

$$E_K \stackrel{\text{def}}{=} \left\{ \hat{W}(\hat{\Theta}_K) \neq W \right\}$$

The error probability of each of the interlaced schemes is  $p_e = \Pr(E_K)$ , and hence the total error probability is upper bounded by  $2p_e$ . Therefore, below we analyze only a single scheme, since the 2 factor does not change the exponential behavior. As in the analysis of the scalar scheme, the channel  $\Theta \rightarrow \mathbf{Y}_K$  is not Gaussian due to the non-linear modulo operations, which complicates a direct analysis. In order to circumvent this, we will upper bound the error probability by further taking modulo-error events into account, and working in the coupled system as before. These error events are defined by

$$E_k \stackrel{\text{def}}{=} \left\{ \gamma_k \epsilon_k + \tilde{\mathbf{Z}}_k \notin \mathcal{V}_0 \right\}$$

and we have that

$$p_e \leq \Pr \left( \bigcup_{k=1}^K E_k \right).$$

Applying the coupling argument of Lemma 1, we can obtain

$$\Pr \left( \bigcup_{k=1}^K E_k \right) = \Pr \left( \bigcup_{k=1}^K E'_k \right).$$

Using the union bound we obtain

$$p_e \leq \sum_{k=1}^K \Pr(E'_k).$$

Calculating the above probabilities now involves only Gaussian random vectors, which significantly simplifies the analysis.

From this point on, we perform an asymptotic exponential analysis. We set the parameters such that all modulo-aliasing error probabilities are equal. Hence

$$p_e \leq \sum_{k=1}^K \Pr(E'_k) \leq \max(p_{\text{mod}}, p_{\text{dec}})$$

where without loss of asymptotic optimality we have set all the modulo-error probabilities to be equal  $p_{\text{mod}} \stackrel{\text{def}}{=} \Pr(E'_k)$ , and also defined  $p_{\text{dec}} \stackrel{\text{def}}{=} \Pr(E'_K)$ . The modulo-aliasing error can be exponentially upper bounded by the Poltyrev exponent (38), i.e.  $p_{\text{mod}} \leq e^{-N_\Lambda E_p(L)}$ .

We observe that the channel  $\Theta \rightarrow \hat{\Theta}_K$  is in fact equivalent (in the coupled system) to  $N_\Lambda$  parallel independent AWGN channels each with the same noise variance  $\sigma_K^2$ , and with a signal-to-noise ratio  $\text{SNR}_K(L)$  given in (39). We can now encode the message  $W$  into  $\Theta$  using a Gaussian codebook of block length  $N_\Lambda$  and rate  $KR$  to obtain

$$p_{\text{dec}} \leq e^{-N_\Lambda E_r(\text{SNR}_K(L), KR)}.$$

Note that the rate  $KR$  is chosen such that the overall rate over  $K$  rounds is  $R$ . Balancing the exponents for  $p_{\text{dec}}$  and  $p_{\text{mod}}$  yields the result. The division by  $2K$  is due to the use of two interlaced schemes, that doubles the overall delay.

The trade-off is now clear: setting the lattice looseness  $L$  to be large reduces  $p_{\text{mod}}$  but also reduces  $\text{SNR}_K(L)$  hence increasing  $p_{\text{dec}}$ , and vice versa. Due to the monotonicity of  $E_r(\text{SNR}_K(L), KR), E_p(L)$  in  $L$ , a numerical solution to (40) can be easily found.

*Remark 7:* It should be noted that as in the simple interaction case, the feedforward transmission power should be reduced by a suitable factor  $\Psi_3$  in order to avoid excess power due to modulo-aliasing errors. However, for high-dimensional lattices, the probability of a modulo-aliasing error is exponentially small in the dimension of the lattice, and the magnitude of the vectors in the fundamental Voronoi cell  $\mathcal{V}_0$  is upper

bounded by  $P$  [22, Section VIII]. Therefore, the power back-off term  $\Psi_3$  is asymptotically negligible and does not affect the statement of the theorem.

#### F. Concatenated Coding with Simple Interaction

In this subsection we introduce a concatenated coding scheme based on the simple interaction scheme of Section V used in conjunction with a (non-interactive) block code. The simple interaction scheme is used as an inner code, inducing a DMC whose input and output correspond to a PAM constellation, and a suitable block code is employed as an outer code over that induced DMC. If the transition matrix pertaining to the induced DMC had been known, the calculation of the error exponent would have been straightforward. Unfortunately, computing this transition matrix appears to be extremely involved. Nevertheless, there is one property of the transition matrix that is already available to us from the preceding discussion - an upper bound on the error probability, namely the probability that the output differs from the input. Therefore, to circumvent the difficulty in computing the error exponent of the induced DMC, we bound it from below by computing the error exponent corresponding to a “worst case” symmetric DMC, that has an error probability equal to (the upper bound on) the error probability induced by our inner coding scheme. This leads to the following result:

*Theorem 3:* The concatenated coding scheme described above, with total delay of  $N$  time instants, attains an error probability  $p_e \leq e^{-N E_{\text{FB}}^{\text{DMC}}(R)}$ , where

$$E_{\text{FB}}^{\text{DMC}}(R) \stackrel{\text{def}}{=} \max_{\delta \in (0, \frac{1}{2}), K \in \mathbb{N}} \left\{ \frac{E^{\text{DMC}}(\delta, M, KR)}{K} \right\}. \quad (41)$$

$M = \left\lfloor \sqrt{1 + \frac{\text{SNR}_K}{\Gamma_0(\frac{\delta}{2})}} \right\rfloor$  is the size of the PAM constellation,  $\text{SNR}_K$  is given in (32) and

$$E^{\text{DMC}}(\delta, M, R) = \max \{ E_r^{\text{DMC}}(\delta, M, R), E_{\text{ex}}^{\text{DMC}}(\delta, M, R) \}.$$

The random coding error exponent is given by

$$E_r^{\text{DMC}}(\delta, M, R) = \max_{0 \leq \rho \leq 1} \left\{ -\rho \ln(2)R + \rho \ln(M) - (1 + \rho) \ln \left( (1 - \delta)^{\frac{1}{1+\rho}} + (M - 1)^{\frac{\rho}{1+\rho}} \delta^{\frac{1}{1+\rho}} \right) \right\} \quad (42)$$

and the expurgated error exponent is given by

$$E_{\text{ex}}^{\text{DMC}}(\delta, M, R) = \sup_{\rho \geq 1} \left\{ -\rho \ln(2)R + \rho \ln(M) - \rho \ln \left( 1 + (M - 1) \left( 2\sqrt{\frac{\delta(1-\delta)}{M-1}} + \frac{M-2}{M-1} \delta \right)^{\frac{1}{\rho}} \right) \right\}. \quad (43)$$

Let  $P$  denote an  $M \times M$  transition matrix of a DMC, where  $P(i | j)$  is the probability that the output is  $j$  given that the input is  $i$ . The maximal error probability of  $P$  is defined as

$$\text{maxerr}(P) \stackrel{\text{def}}{=} 1 - \min_i P(i | i)$$

We say that  $P$  is *totally symmetric*, if

$$P(i | j) = \begin{cases} 1 - \delta & \text{if } i = j \\ \frac{\delta}{M-1} & \text{if } i \neq j \end{cases}.$$

For some  $\delta$ . Clearly, for such  $P$  it holds that  $\text{maxerr}(P) = \delta$ .

The proof of the theorem is based on the observation that a totally symmetric DMC has the smallest uniform input random coding / expurgated error exponent for a given maximal error probability, together with straightforward calculations of the classic error exponents [6] for this channel. Specifically, denote the random coding error exponent at rate  $R$  for a DMC with transition matrix  $P$  and uniform input distribution by  $E_r(R, P)$ . The corresponding expurgated error exponent is denoted by  $E_{\text{ex}}(R, P)$ . The worst case property of the totally symmetric DMC is stated in the following lemma, which is proved in Appendix A:

*Lemma 2:* Let  $P$  be a  $M \times M$  DMC with  $\text{maxerr}(P) \leq \delta$ , and let  $P_{\text{sym}}$  be a totally symmetric  $M \times M$  DMC with  $\text{maxerr}(P_{\text{sym}}) = \delta$ . Then  $E_r(R, P_{\text{sym}}) \leq E_r(R, P)$  and  $E_{\text{ex}}(R, P_{\text{sym}}) \leq E_{\text{ex}}(R, P)$ .

*Proof of Theorem 3:* Let  $P$  be the transition matrix pertaining to the  $M \times M$  DMC induced by our simple interaction scheme, with some target error probability  $\delta$  and  $K$  interaction rounds (the explicit derivation of the scheme parameters and the error analysis are given in Section V). This DMC satisfies  $\text{maxerr}(P) \leq \delta$ . Let  $P_{\text{sym}}$  denote the transition matrix of a totally symmetric  $M \times M$  DMC with  $\text{maxerr}(P_{\text{sym}}) = \delta$ . Lemma 2 implies that the error exponent attained by an optimal outer block code is lower bounded by the random coding / expurgated exponents of  $P_{\text{sym}}$ , where the latter follow from a straightforward calculation and are given by (42) and (43). Finally, (41) follows from the appropriate normalization in the inner code delay  $K$ , and by an optimization over the target error probability and the number of interaction rounds. ■

#### G. Discussion

Numerical evaluations of  $E_{\text{sp}}$ ,  $E_r$ ,  $E_{\text{FB}}$  and  $E_{\text{FB}}^{\text{DMC}}$  for  $\text{SNR} = 20\text{dB}$  are depicted in Fig. 6. It can be clearly seen that for low rates  $E_{\text{FB}}^{\text{DMC}}$  is larger than  $E_{\text{FB}}$  for both value of  $\Delta\text{SNR}$ . What is less visible in Fig. 6 but is clearer in Fig. 7 is that at high rates  $E_{\text{FB}}$  exceeds  $E_{\text{FB}}^{\text{DMC}}$ . Comparing these achievable exponent to the sphere packing bound, which upper bounds the best achievable error exponent without feedback, it is evident that both error exponents are smaller than the sphere packing



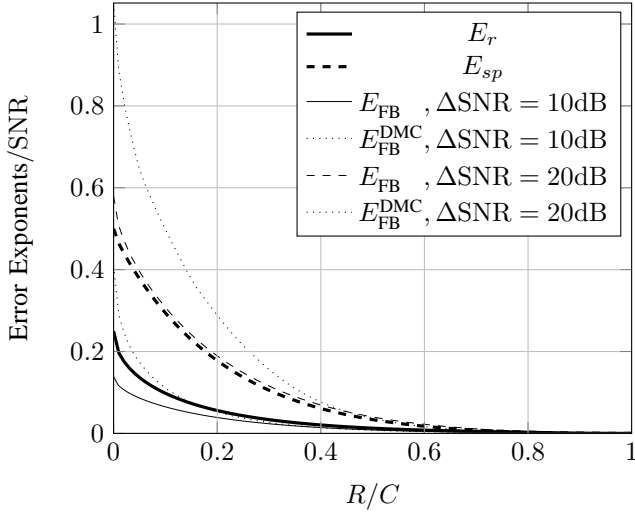


Fig. 6. Error exponents with and without feedback for SNR = 20dB with  $\Delta\text{SNR} = 10\text{dB}$  and  $\Delta\text{SNR} = 20\text{dB}$

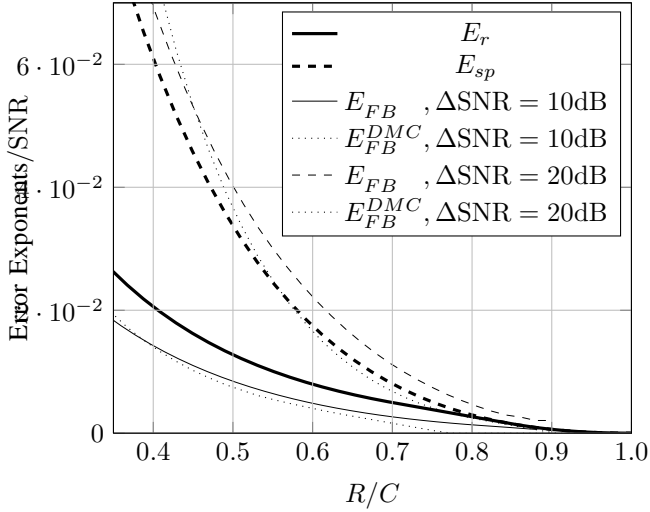


Fig. 7. Fig. 6 at high rates

bound for  $\Delta\text{SNR} = 10\text{dB}$ , and the maximal between them is greater than the sphere packing bound for  $\Delta\text{SNR} = 20\text{dB}$ .

It is of interest to compare the error exponents of Theorems 2 and 3 to the noisy feedback error exponent of [12], which employs a concatenated linear coding scheme. In this scheme the output of the inner code is a linear combination of the message variable and noisy versions of the previous channel outputs (i.e. *passive* feedback), and the outer code is a random block code. Due to the linear nature of the inner code, its output is regarded as a single use of an AWGN channel, and the scheme parameters are judiciously chosen to try and maximize its SNR. Then, the error exponent of the concatenated scheme is readily calculated using the standard AWGN error exponents.

Fig. 8 compares our error exponents to that of [12], which is denoted by  $E_{\text{FB}}^{CL}$ . In this setting SNR = 10dB and  $\Delta\text{SNR} =$

23dB. We can see that  $E_{\text{FB}}^{CL}$  dramatically improves over the non-feedback achievable error exponent at rates close to zero, but then falls below it at rates above  $0.46C$  (where  $C$  is the channel capacity). Comparing  $E_{\text{FB}}^{CL}$  to  $E_{\text{FB}}^{\text{DMC}}$  we can see that the latter is better at rates higher than  $0.18C$ . Finally,  $E_{\text{FB}}$  is the best among the three feedback error exponent at rates above  $0.53C$ , and also beats the sphere packing bound at rates up to  $0.9C$ .

It is interesting to note that in [12], the authors characterize the “shut-off rate”  $R_{th}^{CL}$  of their scheme, namely as the rate above which their exponent does not improve the no-feedback one. This rate is given by [12, Lemma 8]:

$$R_{th}^{CL} = \frac{1}{4} \log \left( 1 + 2\text{SNR} \cdot \Delta\text{SNR} \cdot \frac{\text{SNR}}{1 + \text{SNR}} (1 - \gamma_0) \right)$$

where  $\gamma_0 \in [0, 1]$  is a root of a quadratic equation given in their Lemma 6 (note that a factor  $\frac{1}{2}$  is missing in the original expression for  $R_{th}^{CL}$ ). Dividing the above by the capacity, it is easy to show that

$$\frac{R_{th}^{CL}}{C} \leq \frac{1}{2} \left( 1 + \frac{1 + \log(1 + \Delta\text{SNR})}{\log \text{SNR}} + \frac{(1 + \text{SNR}) \log e}{2\Delta\text{SNR} \cdot \text{SNR}^2 \log \text{SNR}} \right)$$

For  $\text{SNR} \gg 1$  and a fixed  $\Delta\text{SNR}$ , the above upper bound clearly converges to  $\frac{1}{2}$ , hence in this high-SNR regime the error exponent of [12] does not yield any improvement over the non-feedback exponent for rates above half the capacity.

It is also instructive to compare the error exponents discussed above at zero rate. The zero-rate exponent attained by the scheme in [12] is given in Lemma 8 therein:

$$E_{\text{FB}}^{CL}(R = 0) = \frac{\text{SNR}}{4} \left( 1 + \Delta\text{SNR} \cdot \frac{\text{SNR}}{1 + \text{SNR}} \right) \quad (44)$$

It can be verified by direct calculation that at zero rate, the error exponent of our concatenated scheme outperforms that of the lattice-based scheme. The former is given in the following formula at high-SNR, derived in Appendix B:

$$E_{\text{FB}}^{\text{DMC}}(R = 0) \geq \frac{\frac{3}{2}\Delta\text{SNR} (e^{-1}\text{SNR} - 1) - \ln(2\lfloor \ln(3\Delta\text{SNR}) \rfloor - 1)}{4\lfloor \ln(3\Delta\text{SNR}) \rfloor} \quad (45)$$

It is easy to see that in the limit of  $\text{SNR} \rightarrow \infty$  (44) is larger than (45).

## VII. ACKNOWLEDGMENTS

We thank the anonymous reviewers for many helpful comments and observations that have significantly improved the presentation of the paper.

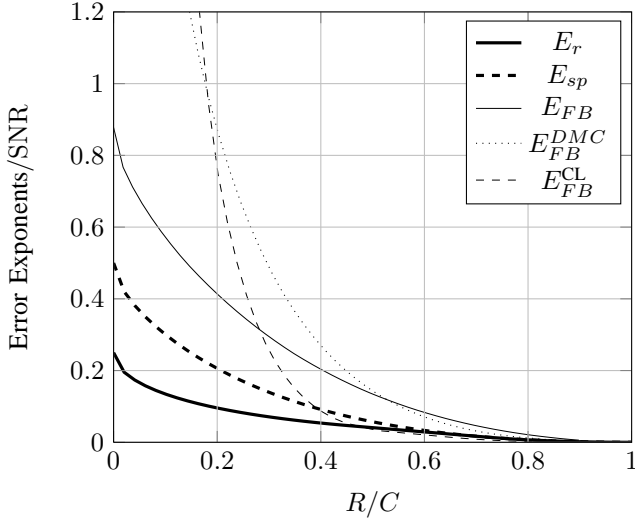


Fig. 8. A comparison of non-feedback error exponents, our error exponents ( $E_{FB}$ ,  $E_{FB}^{DMC}$ ) and Chance and Love's error exponent ( $E_{FB}^{CL}$ ). In this setting SNR = 10dB and  $\Delta\text{SNR} = 23\text{dB}$

#### APPENDIX A PROOF LEMMA 2

The proof is based on a symmetrization argument in conjunction with a convexity argument. We begin by recalling the error exponents for a DMC with transition matrix  $P$ , input distribution  $Q$  and rate  $R$  [6]. The random coding exponent is given by

$$E_r(R, Q, P) = \max_{0 \leq \rho \leq 1} \left\{ -\rho \ln(2)R - \ln \sum_{j=0}^{M-1} \left[ \sum_{k=0}^{M-1} Q(k) P(j|k)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\},$$

and the expurgated error exponent is given by

$$E_{ex}(R, Q, P) = \max_{\rho \geq 1} \left\{ -\rho \ln(2)R - \rho \ln \sum_{k,i} Q(k)Q(i) \left[ \sum_j \sqrt{P(j|k)P(j|i)} \right]^{\frac{1}{\rho}} \right\}.$$

Below we restrict our attention to a uniform input distribution  $Q(i) = \frac{1}{M}$  for  $0 \leq i \leq M-1$ , which can be suboptimal but facilitates a much simpler analysis. We denote the random coding and expurgated exponents corresponding to a uniform input by  $E_r(R, P)$  and  $E_{ex}(R, P)$  respectively. These are explicitly given by

$$E_r(R, P) = \max_{0 \leq \rho \leq 1} \{ -\rho \ln(2)R + (1+\rho) \ln(M) - \ln F_r(\rho, P) \}, \quad (46)$$

and

$$E_{ex}(R, P) = \max_{\rho \geq 1} \{ -\rho \ln(2)R + 2\rho \ln(M) - \rho \ln F_{ex}(\rho, P) \}, \quad (47)$$

where

$$F_r(\rho, P) \stackrel{\text{def}}{=} \sum_{j=0}^{M-1} \left[ \sum_{k=0}^{M-1} P(j|k)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

and

$$F_{ex}(\rho, P) \stackrel{\text{def}}{=} \sum_{k,i} \left[ \sum_j \sqrt{P(j|k)P(j|i)} \right]^{\frac{1}{\rho}}.$$

Now, let  $\Pi$  denote the set of all possible permutations over  $\{1, \dots, M\}$ . For a given channel transition matrix  $P$ , and a given permutation  $\pi \in \Pi$ , we define the channel transition matrix  $P_\pi$  to be

$$P_\pi(j|i) \stackrel{\text{def}}{=} P(\pi(j)|\pi(i)).$$

The following lemma easily follows from (46) and (47) by incorporating the permutation into the order of summations.

**Lemma 3:**  $E_r(R, P_\pi) = E_r(R, P)$  and  $E_{ex}(R, P_\pi) = E_{ex}(R, P)$  for any  $\pi \in \Pi$ .

We further define the symmetrized channel matrix  $P_{sym}$  associated with  $P$ , to be the one obtained by uniformly averaging over all possible permutations, i.e.,

$$P_{sym}(j|i) \stackrel{\text{def}}{=} \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P_\pi(j|i).$$

We now have the following lemma.

**Lemma 4:**  $P_{sym}$  is a totally symmetric DMC with  $\text{maxerr}(P_{sym}) \leq \text{maxerr}(P)$ .

*Proof:* Choose some indices  $i, j, k, \ell$ , such that if  $i = j$  then also  $k = \ell$ . Let  $\sigma \in \Pi$  be the permutation where  $\sigma(i) = k$ ,  $\sigma(j) = \ell$ , and  $\sigma(x) = x$  otherwise.

$$\begin{aligned} P_{sym}(j|i) &= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P(\pi(j)|\pi(i)) \\ &= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P((\pi \circ \sigma)(j)|(\pi \circ \sigma)(i)) \\ &= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P(\pi(k)|\pi(\ell)) \\ &= P_{sym}(k|\ell). \end{aligned}$$

Hence  $P_{sym}$  is a totally symmetric DMC. Specifically, for any  $i$

$$\begin{aligned} P_{sym}(i|i) &= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P(\pi(i)|\pi(i)) \\ &= \frac{1}{|\Pi|} \sum_k \sum_{\pi \in \Pi, \pi(i)=k} P(k|k) \\ &= \frac{1}{M} \sum_k P(k|k). \end{aligned}$$

Thus,

$$\begin{aligned}\maxerr(P_{sym}) &= 1 - \frac{1}{M} \sum_k P(k|k) \\ &\leq 1 - \min_k P(k|k) \\ &= \maxerr(P).\end{aligned}$$

■

We now proceed to show that the (uniform input) error exponents associated with  $P_{sym}$  can be used as lower bounds for those of  $P$ . This follows from a concavity argument combined with Lemma 3.

*Lemma 5:*  $E_r(R, P_{sym}) \leq E_r(R, P)$  and  $E_{ex}(R, P_{sym}) \leq E_{ex}(R, P)$ .

*Proof:* We begin by showing that both  $F_r(\rho, P)$  (in (46)) and  $F_{ex}(\rho, P)$  (in (47)) are concave in  $P$  for all valid values of  $\rho$ . We start with  $F_r(\rho, P)$  and recall its definition:

$$F_r(\rho, P) = \sum_{j=0}^{M-1} \left[ \sum_{k=0}^{M-1} P(j|k)^{\frac{1}{1+\rho}} \right]^{1+\rho}.$$

We note that  $\sum_{k=0}^{M-1} P(j|k)^{\frac{1}{1+\rho}}$  is a concave function in the vector  $P(j|\cdot)$ ; this follows by observing that this function is a norm of order  $\frac{1}{1+\rho} \in [\frac{1}{2}, 1]$  over  $\mathbb{R}_+^M$  (and the values of  $P$  are strictly positive by design).  $F_r(\rho, P)$  is thus concave in  $P$  as well, being a sum of concave functions.

Similarly, we recall the definition of  $F_{ex}(\rho, P)$

$$F_{ex}(\rho, P) \stackrel{\text{def}}{=} \sum_{k,i} \left[ \sum_j \sqrt{P(j|k)P(j|i)} \right]^{\frac{1}{\rho}}.$$

We note that  $\sqrt{P(j|k)P(j|i)}$  is a geometric mean, which is concave in the vector  $(P(j|k), P(j|i)) \in \mathbb{R}_+^M$  for any fixed  $k$ . Thus,  $\sum_j \sqrt{P(j|k)P(j|i)}$  is concave as well, being a sum of concave functions. The  $\frac{1}{\rho}$  power of this sum is concave since it is a scalar composition of a concave function with a concave nondecreasing function  $x^{\frac{1}{\rho}}$  (for  $\rho \geq 1$ ). Finally,  $F_{ex}(\rho, P)$  is a sum of concave functions hence is also concave.

With concavity in hand, we can readily apply Jensen's inequality

$$\begin{aligned}F_r(\rho, P_{sym}) &= F_r \left( \rho, \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P_\pi \right) \\ &\geq \frac{1}{|\Pi|} \sum_{\pi \in \Pi} F_r(\rho, P_\pi) \\ &= F_r(\rho, P)\end{aligned}\tag{48}$$

$$= F_r(\rho, P)\tag{49}$$

where (48) follows from concavity via Jensen's inequality, and (49) is by virtue of Lemma. 3. Plugging the resulting inequality in (46) yields  $E_r(R, P_{sym}) \leq E_r(R, P)$ . Following the exact same steps for  $F_{ex}$  we can prove that  $E_{ex}(R, P_{sym}) \leq E_{ex}(R, P)$ . ■

Let us now calculate the error exponents for  $P_{sym}$ . Since by Lemma 3 we have that  $P_{sym}$  is a totally symmetric DMC, and denoting  $\delta \stackrel{\text{def}}{=} \maxerr(P_{sym})$ , we can use (46) and (47) to obtain

$$E_r(R, P_{sym}) = \max_{0 \leq \rho \leq 1} \left\{ -\rho \ln(2)R + \rho \ln(M) - (1+\rho) \ln \left( (1-\delta)^{\frac{1}{1+\rho}} + (M-1)^{\frac{\rho}{1+\rho}} \delta^{\frac{1}{1+\rho}} \right) \right\}\tag{50}$$

and

$$E_{ex}(R, P_{sym}) = \sup_{\rho \geq 1} \left\{ -\rho \ln(2)R + \rho \ln(M) - \rho \ln \left( 1 + (M-1) \left( 2\sqrt{\frac{\delta(1-\delta)}{M-1}} + \frac{M-2}{M-1} \delta \right)^{\frac{1}{\rho}} \right) \right\}.\tag{51}$$

Lemma 3 also tells us that  $\delta \leq \maxerr P$ . It can be verified (by direct differentiation) that both (50) and (51) are monotonically decreasing in  $\delta$  for  $\delta \in (0, \frac{1}{2})$ . Therefore, replacing  $\delta$  with any upper bound on  $\maxerr(P)$  still results in a lower bound for both  $E_r(R, P_{sym})$  and  $E_{ex}(R, P_{sym})$ , concluding the proof.

## APPENDIX B

### ZERO RATE ANALYSIS FOR THEOREM 3

Below we provide a lower bound for the error exponent of Theorem 3 at  $R = 0$ . We use the expurgated error exponent which is known to be larger at  $R = 0$ , and use  $M = 2$  in (43) which yields

$$E_{ex}^{\text{DMC}}(\delta, M = 2, R = 0) = \sup_{\rho \geq 1} \left\{ \rho \ln(2) - \rho \ln \left( 1 + \left( 2\sqrt{\delta(1-\delta)} \right)^{\frac{1}{\rho}} \right) \right\}.$$

It can be shown (see [6, Problem 5.24]) that for every input distribution, the zero rate error exponent is optimized by taking the limit  $\rho \rightarrow \infty$ , which in the case of uniform input results in:

$$\begin{aligned}E_{ex}^{\text{DMC}}(\delta, M = 2, R = 0) &= -\frac{1}{2} \ln \left( 2\sqrt{\delta(1-\delta)} \right) \\ &= -\frac{1}{4} \ln(2\delta(1-\delta)) \\ &\geq -\frac{1}{4} \ln(2\delta)\end{aligned}\tag{52}$$

Let us now find a simple setting for the scheme's parameters. Applying the union bound it is clear that the error probability of the inner code is upper bounded by  $\delta$ , where

$$\delta = Q \left( \sqrt{\text{SNR}_K} \right) + 2(K-1)Q \left( \sqrt{3L} \right),\tag{53}$$

where the first addend is due to the probability of error at the PAM (here, BPSK) decoder, and the second addend is due to the modulo-aliasing error probabilities. In order to simplify the solution, we impose equality between the two  $Q$ -functions,

which results in  $\text{SNR}_K = 3L$ . Using (32) and some algebra, we obtain

$$\text{SNR} \left( \frac{\Delta \text{SNR}(1 + \text{SNR})}{\Delta \text{SNR} + L} \right)^{K-1} = 3L. \quad (54)$$

This equation is difficult to solve. To simplify, we note that solution of the following equation for  $L$ :

$$\text{SNR} \left( \frac{\Delta \text{SNR}(1 + \text{SNR})}{\Delta \text{SNR} + L} \right)^{K-1} = 3(\Delta \text{SNR} + L),$$

gives a lower bound on the corresponding solution of (54), yielding

$$L = -\Delta \text{SNR} + 3^{-\frac{1}{K}} \Delta \text{SNR}^{1-\frac{1}{K}} (\text{SNR}(1 + \text{SNR})^{K-1})^{\frac{1}{K}} \quad (55)$$

It is instructive to use a high-SNR approximation  $1 + \text{SNR} \approx \text{SNR}$  and compute an approximate (yet attainable) version of (55):

$$L = \Delta \text{SNR} \left( (3\Delta \text{SNR})^{-\frac{1}{K}} \text{SNR} - 1 \right)$$

Plugging this into (53) and using the exponential bound for the Q-function:  $Q(x) \leq \frac{1}{2} \exp(-\frac{1}{2}x^2)$ , we get

$$\delta \leq (2K - 1)Q(\sqrt{3L}) \leq \frac{2K - 1}{2} \exp\left(-\frac{3}{2}L\right)$$

Now, plugging the above into the error exponent bound (52) and recalling the concatenated coding exponent (41), results in

$$\begin{aligned} E_{\text{FB}}^{\text{DMC}}(R = 0) &\geq \max_{K \in \mathbb{N}} \left\{ \frac{\frac{3}{2}L - \ln(2K - 1)}{4K} \right\} \\ &= \max_{K \in \mathbb{N}} \left\{ \frac{\frac{3}{2}\Delta \text{SNR} \left( (3\Delta \text{SNR})^{-\frac{1}{K}} \text{SNR} - 1 \right) - \ln(2K - 1)}{4K} \right\} \end{aligned}$$

Let us now approximately optimize for  $K$ , by taking the derivative to zero in a simplified expression

$$\frac{(3\Delta \text{SNR})^{-\frac{1}{K}}}{K}$$

yielding  $K^* = \lfloor \ln(3\Delta \text{SNR}) \rfloor$  and  $(3\Delta \text{SNR})^{-\frac{1}{K^*}} \geq (3\Delta \text{SNR})^{-\frac{1}{\ln(3\Delta \text{SNR})}} = e^{-1}$  finally obtaining the following achievable error exponent:

$$\begin{aligned} E_{\text{FB}}^{\text{DMC}}(R = 0) &\geq \\ &\frac{\frac{3}{2}\Delta \text{SNR} (e^{-1}\text{SNR} - 1) - \ln(2\lfloor \ln(3\Delta \text{SNR}) \rfloor - 1)}{4\lfloor \ln(3\Delta \text{SNR}) \rfloor} \end{aligned} \quad (56)$$

Let us now evaluate  $E_{\text{FB}}^{\text{DMC}}(R = 0)$  for  $\text{SNR} \gg 1$  when  $\Delta \text{SNR}$  is held fixed. In this case (56) can be approximately expressed as

$$E_{\text{FB}}^{\text{DMC}}(R = 0) \gtrsim \frac{3\Delta \text{SNR}}{8e\lfloor \ln(3\Delta \text{SNR}) \rfloor} \text{SNR}. \quad (57)$$

Note that by (37) the sphere packing bound at  $R = 0$  is  $\frac{1}{2}\text{SNR}$  rendering (57) better for  $\Delta \text{SNR}_{\text{dB}} \gtrsim 10.4\text{dB}$ . It is interesting to note that in this setting (44) still yields a larger value.

## REFERENCES

- [1] C. E. Shannon, "The zero-error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. IT-2, pp. 8–19, Sep 1956.
- [2] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback part I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 172–182, Apr 1966.
- [3] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback part II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 183–189, Apr 1966.
- [4] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Info. Theory*, vol. IT-9, pp. 136–143, Jul 1963.
- [5] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, Mar 2011.
- [6] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968.
- [7] Y.-H. Kim, A. Lapidoth, and T. Weissman, "On the reliability of Gaussian channels with noisy feedback," in *Proc. 41st Allerton Conf. Communication, Control Computing*, Sep. 2006, pp. 364–371.
- [8] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 6–17, 2010.
- [9] Y. Kochman and R. Zamir, "Joint Wyner-Ziv/Dirty-Paper Coding by Analog Modulo-Lattice Modulation," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4878–4889, 2009.
- [10] M. V. Burnashev and H. Yamamoto, "Noisy feedback improves the gaussian channel reliability function," in *ISIT*, 2014, pp. 2554–2558.
- [11] Y. Xiang and Y.-H. Kim, "Gaussian channel with noisy feedback and peak energy constraint," *Information Theory, IEEE Transactions on*, vol. 59, no. 8, pp. 4746–4756, 2013.
- [12] Z. Chance and D. J. Love, "Concatenated coding for the AWGN channel with noisy feedback," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6633–6649, Oct. 2011.
- [13] Y.-H. Kim, A. Lapidoth, and T. Weissman, "Error exponents for the Gaussian channel with active noisy feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1223–1236, March 2011.
- [14] A. Ben-Yishai and O. Shayevitz, "The Gaussian Channel with Noisy Feedback: Improving Reliability via Interaction," in *ISIT 2015*, 2015, pp. 2500–2504, Also available in <http://arxiv.org/abs/1501.06671>.
- [15] A. Ben-Yishai and O. Shayevitz, "The Gaussian Channel with Noisy Feedback: Near-Capacity Performance via Simple Interaction," in *Proc. 52nd Allerton Conf. Communication, Control Computing*, Oct. 2014, pp. 152–159, Also available in <http://arxiv.org/abs/1407.8022>.
- [16] M. V. Burnashev, "Data transmission over a discrete channel with feedback," *Probl. Pered. Inf.*, vol. 12, no. 4, pp. 10–30, 1976, English translation in *Probl. Inf. Transm.*, pp. 250–265, 1976.
- [17] A. Sato and H. Yamamoto, "Error exponents of discrete memoryless channels and awgn channels with noisy feedback," in *ISITA*, 2010, pp. 452–457.
- [18] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 558–567, Oct 1965.
- [19] A. El Gamal and Y.-H. Kim, *Network information theory*, Cambridge University Press, 2011.
- [20] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.
- [21] R. Zamir, *Lattice Coding for Signals and Networks*, Cambridge University Press, 2014.

- [22] U. Erez, S. Litsyn and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct 2005.

**Assaf Ben-Yishai** received the B.Sc. degree (summa cum laude) and M.Sc. degree (magna cum laude) in 1999 and 2001 respectively, both in electrical engineering, and both from the Tel-Aviv university. Assaf worked for several years in leading algorithmic positions in the Israeli high-tech industry, primarily in the design of advanced features of wireless and wire-line communication systems. Currently he is pursuing his Ph.D. degree in the field of information theory.

**Ofer Shayevitz** received the B.Sc. degree (summa cum laude) from the Technion Institute of Technology, Haifa, Israel, in 1997 and the M.Sc. and Ph.D. degrees from the Tel-Aviv University, Tel Aviv, Israel, in 2004 and 2009, respectively, all in electrical engineering. He is currently a Senior Lecturer in the Department of EE - Systems at the Tel Aviv University, and also serves as the head of the Advanced Communication Center (ACC). Before joining the department, he was a postdoctoral fellow in the Information Theory and Applications (ITA) Center at the University of California, San Diego, from 2008 to 2011, and worked as a quantitative analyst with the D. E. Shaw group in New York from 2011 to 2013. Prior to his graduate studies, he served as an engineer and team leader in the Israeli Defense Forces from 1997 to 2003, and as an algorithms engineer at CellGuide from 2003 to 2004. Dr. Shayevitz is the recipient of the ITA postdoctoral fellowship (2009 - 2011), the Adams fellowship awarded by the Israel Academy of Sciences and Humanities (2006 - 2008), the Advanced Communication Center (ACC) Feder Family award (2009), and the Weinstein prize (2006 - 2009).